Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Visual Reaction Time and the Broca-Sulzer Phenomenon: D. RAAB, E. FEHRER, AND M. HERSHENSON 193

The Organization of Component Response Error Events in Two-Dimensional Visual Tracking: J. A. ADAMS AND C. E. WEBBER 200

The Effects of Anxiety Level and Shock on a Paired-Associate Verbal Task: L. C. LEE 213

The Probability of Detecting Small Numbers of Dots: F. RESTLE, J. RAE, AND C. KIESLER 218

Subjective Similarity in Two Perceptual Continua: G. EKMAN, G. GOUDE, AND Y. WAERN 222

Studies of Distributed Practice: XX. Sources of Interference Associated with Differences in Learning and Retention: B. J. UNDERWOOD AND R. W. SCHULZ 228

Studies in Incidental Learning: IX. A Comparison of the Methods of Successive and Single Recalls: L. POSTMAN AND L. W. PHILLIPS 236

The Effect of Speed-Up Instructions on Spatial Stimulus Generalization: B. W. HARLESTON 242

Response Speed, Amplitude, and Resistance to Extinction as Joint Functions of Work and Length of Behavior Chain: R. F. WEISS 245

Supplementary Report: Differential Position Habits and Anxiety in Children as Determinants of Performance in Learning: A. CASTANEDA 257

Supplementary Report: Effect of Redundant Relevant Information upon the Identification of Concepts: L. E. BOURNE, JR., AND R. C. HAYGOOD 259

Supplementary Report: The Effects of the Mean, Midpoint, and Median upon Adaptation Level in Judgment: A PARDUCCI AND L. M. MARSHALL 261

Supplementary Report: Recognition Responses to Ambiguous Cues Following Paired-Associate Learning: A. BINDER 263

American Psychological Association

Vol. 61, No. 3 March 1961

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 61, NO. 3 MARCH 1961

VISUAL REACTION TIME AND THE BROCA–SULZER PHENOMENON [1]

DAVID RAAB, ELIZABETH FEHRER, AND MAURICE HERSHENSON

Brooklyn College

If a brief light flash is followed by additional visual stimulation, it may appear less bright than if it had been flashed alone. The Broca-Sulzer phenomenon, metacontrast (e.g., Alpern, 1953; Stigler, 1910) and the Crawford effect (Crawford, 1947) are examples of such retroactive masking.

A theoretically important question that has received little systematic study concerns the behavioral effectiveness of a stimulus thus masked. Will overt responses to it be similar to the responses evoked by a less intense stimulus which phenomenally matches the masked one? This question was raised in 1942 by Baumgardt and Segal and led to an experiment in which reaction time (RT) was used as a measure of the Broca-Sulzer phenomenon.

This phenomenon, which has had ample verification, consists in the fact that the phenomenal brightness of a stimulus can be diminished by the unchanged continuation of the stimulus itself. Studies agree (e.g., Bills, 1920; Broca & Sulzer, 1902; Stainton, 1928) that the apparent brightness of a flash of light increases with stimulus duration up to 50 to 100 msec., then decreases up to about 250 msec., after which flash duration has a comparatively negligible influence on apparent brightness. The overshooting effect is most marked with intense stimuli and is absent with weak ones. The method used in establishing the Broca-Sulzer phenomenon typically consists in matching the short duration test flash to a neighboring or surrounding light presented for several seconds.

The Baumgardt and Segal (1942) experiment was designed to determine whether differences in apparent brightness can be measured by RT. Specifically, they anticipated a shorter RT to a 50-msec. flash than to a flash of 250 msec. It should be pointed out that Baumgardt and Segal assumed that the brightness differences that emerge with the Broca-Sulzer method derive purely from differences in stimulus duration. They therefore presented their light flashes unaccompanied by the comparison stimulus employed in the Broca-Sulzer procedure.

The stimulus used had a luminance of 1700 nits, (496 ft-L). The two durations were presented in blocks of 10 trials each, the blocks following each other in ABAB order. Of 31 comparisons between the means of successive blocks, all showed slightly faster (1 to 5 msec.) RTs to the 50-msec. flash. It is interesting to note that the

[1] This research was supported by grants from the National Science Foundation (G-6456) and from the National Institute of Neurological Diseases and Blindness (B-1028) and by funds provided by Brooklyn College. Edward H. Green of the Department of Physics gave generous assistance in the planning of the stimulus-generating equipment.

published RTs were all less than the duration of the 250-msec. stimulus and it is therefore obvious that the RT could not have been affected by that part of the 250-msec. stimulus that persisted after the reaction was completed.

A search of the literature yielded only one other study of the effect of a wide range of stimulus durations on RT to clearly suprathreshold visual stimuli. In 1913, Wells measured RTs to light flashes of the following durations: 12, 25, 64, 144, and 1000 msec. The intensity of the light cannot be determined accurately from the published report but was apparently well under 1 ft-L. Wells' data indicate that, in general, the 25- and 64-msec. flashes yielded the fastest RTs. A series of studies by Chernikoff, Gregg, and Brogden (Chernikoff & Brogden, 1949; Chernikoff, Gregg, & Brogden, 1950; Gregg & Brogden, 1950a, 1950b) on the effect of stimulus duration on RT to auditory stimuli yielded congruent results. RT was found to increase as stimulus duration increased from 100 to 2400 msec.

Neither the Wells nor the Chernikoff et al. data can be explained on the basis of the Broca-Sulzer relation between stimulus duration and apparent brightness. Wells' light intensity was below that at which the Broca-Sulzer phenomenon can be demonstrated and no parallel to this phenomenon has been reported for audition. It seems more likely that the effect of stimulus duration on RT was mediated by attitudinal factors which could develop during a block of trials at the same stimulus duration (cf. Jarl, 1957; Johnson, 1923). Blocks of brief stimuli may arouse *S* to react as fast as possible, while longer lasting ones might encourage some relaxation of effort. This may also be the best explanation of Baumgardt and Segal's data even though their light intensity was sufficiently high to yield the Broca-Sulzer effect.

The present experiments have been designed to determine whether the Baumgardt and Segal findings are better explained on the basis of the Broca-Sulzer phenomenon or on the basis of attitudinal factors. In Exp. I, the stimulus durations were presented in random sequence in order to eliminate the possible effect of attitudinal factors which might develop during blocks of trials at the same duration. Thus, any advantage of the 50-msec. flash could be attributed to the Broca-Sulzer effect. Further, such an outcome would be good evidence that the latency of the response is correlated with perceived intensity, stimulus intensity being held constant. Three luminances, approximately 3000, 30, and 0.30 ft-L, were used. As the Broca-Sulzer phenomenon is apparent only with fairly intense stimuli, it was assumed that, if it does indeed affect RT, its effect should be most obvious with our most intense stimulus and absent with the weakest. Experiment II was designed to test the earlier findings that RT to a 50-msec. flash is faster than RT to a 250-msec. flash when stimuli of a given duration are presented in blocks of trials.

Experiment I

Method

Apparatus.—The light source was a glow modulator tube (Sylvania R-1131C) mounted in a light-tight box with a circular aperture, 1 cm. in diameter, in front of the tube. The light was diffused over a ground glass plate placed 1 cm. in front of the lamp and bearing against the inside surface of the front of the box. All of the transilluminated glass was thus masked off except for the circular area in the aperture. This 1-cm. target, viewed binocularly from a headrest 50 cm. away, subtended 1° 10′ of arc.

The front side of the stimulus box measured 13 by 15.5 cm., and was painted black. The stimulus aperture at the center of this rectangle was surrounded by a ring, .75 cm. wide and 6.3 cm. in outer diameter, cut from a manila folder. This concentric ring served as a fixation device.

The apparatus was located in a sound proof room and was dimly illuminated from behind *S*'s head by a diffuse circular source, 5.5 cm. in diameter, placed 160 cm. from the stimulus area. The luminance of this source was 22 ft-L. When illuminated by this source, the luminance of the black front surface of the stimulus box was too low to be measured by means of a Macbeth illuminometer. The fixation ring had a luminance

of approximately 0.001 ft-L. At all luminances studied, the target flash appeared far brighter than the ring and was the only transient visual stimulus.

The glow modulator tube was connected in series with the plate circuit of a triode-connected 6L6. The triode was biased beyond cutoff by means of a 45-v. battery and was pulsed into conduction by rectangular positive pulses from a set of Tektronix Pulse and Wave-Form generators. The current through the 6L6 (and glow tube) was adjusted to be 40 ma., and was monitored on an oscilloscope by means of a 10 ohm resistor in the cathode circuit of the 6L6. A regulated power supply furnished 350 v. dc.

A 2.7-megohm resistance, connected across the glow lamp, was found to yield optimum switching action. In addition, the driving pulses applied to the grid of the 6L6 were shaped to "over-volt" the glow modulator tube. With these arrangements, the tube could be pulsed once every 13 sec. with latency of 50 μsec. and jitter reduced to less than 10 μsec. Light output from pulse to pulse varied less than 0.5%.

Six generators were used to provide the six stimulus durations. The light pulse amplitudes were further monitored by means of a photocell and were adjusted to equality with an accuracy better than one percent. The six durations were preset and a six-position switch was employed to select each duration as required.

The flash durations used were 10, 25, 50, 100, 250, and 500 msec.

Three luminances covering a range of four log units were used. The highest was approximately 3000 ft-L as measured by a Macbeth illuminometer at the ground glass plate. The next was produced by inserting a Kodak filter, neutral density 2.00, between the tube and the ground glass plate. For the weakest light, an ND 3.00 and an ND 1.00 filter were used. All luminances were clearly suprathreshold.

Each trial began with a 1000-cycle warning tone lasting 1 sec., and was followed by a silent foreperiod of 0.5, 1.0, or 1.5 sec., selected by means of a three-position switch. The light stimulus was presented (and the chronoscope started) at the end of this foreperiod. The whole sequence was presented once every 13 sec. by means of Hunter decade interval timers.

RTs (measured from the onset of the stimulus pulse to *S*'s depressing a normally closed telegraph key) were measured to the nearest millisecond by means of a Berkeley counter.

Subjects.—The three authors served as *S*s. Before the experiment proper began, each had undergone a training period of at least 10 hr.

Procedure.—Each *S* served in 18 experimental sessions. Only one luminance was used in a given session. Each session consisted of four sets of 18 trials each in which each combination of the six durations and the three foreperiods appeared once in random order. Sessions for two *S*s were in the sequence ABCCBACBAABC etc. (A, B, and C refer to the three flash luminances.) For the third *S*, this sequence was reversed. For each luminance-duration combination, there were 72 reactions for each *S*, 24 at each foreperiod.

After 5 min. of dark adaptation, eight practice trials were given before the recorded trials were run. A 2-min. rest period was given between Sets 2 and 3, i.e., after 36 recorded trials.

Results

The mean RT for each *S* for each luminance-duration-foreperiod combination was computed after eliminating the longest and the shortest RT of the 24 at each combination. The means thus computed for each luminance-duration combination are based on 66 RTs, i.e., 72 minus 6. They are presented in Table 1 together with the combined means.

No effect of duration is apparent save possibly for the 10-msec. duration of the weakest light which, on the average, yielded somewhat longer RTs than the other durations at this luminance. Phenomenally, this light appeared quite faint. For all *S*s together, however, an analysis of variance showed that the six duration means did not differ significantly ($P > .05$). With the exception of this 10-msec. light, both individual and combined mean RTs for the six durations at a given luminance were very similar and showed no consistent trends with duration. At the greatest luminance, where the effect of the Broca-Sulzer phenomenon should be

TABLE 1

MEAN REACTION TIMES CLASSIFIED BY FLASH DURATION

Flash Luminance in Ft-L	S	Flash Duration in Milliseconds						All RTs	
		10	25	50	100	250	500	Mean	*SD*
3000	EF	159.4	157.5	158.6	157.9	158.6	157.5	158.2	14.7
	MH	153.4	153.8	155.0	153.6	154.5	156.6	154.5	14.9
	DR	162.9	159.3	161.3	163.3	160.2	160.0	161.2	13.0
	Mean	158.6	156.9	158.3	158.3	157.8	158.0	158.0	14.6
30	EF	171.8	169.4	170.9	170.9	175.2	171.6	171.6	16.2
	MH	165.3	165.5	166.6	166.9	165.2	165.6	165.8	15.4
	DR	171.9	171.1	171.3	170.4	171.0	170.7	171.1	12.2
	Mean	169.7	168.7	169.6	169.4	170.4	169.3	169.5	15.1
0.30	EF	201.4	199.2	195.9	197.2	199.7	194.3	198.0	17.1
	MH	195.3	188.4	188.8	191.0	186.4	195.7	190.9	18.5
	DR	206.0	201.1	203.6	201.7	202.8	203.9	203.2	14.6
	Mean	200.9	196.2	196.1	196.6	196.3	198.0	197.4	18.0

Note.—Each individual mean is based on 66 measures.

most apparent, the RTs to the 50- and to the 250-msec. flashes were almost identical for each *S*. With the 3000- and the 30-ft-L lights, the 25-msec. flash yielded (for the three *S*s combined) a somewhat faster RT than any of the others, but the difference between the RT to the 25-msec. flash and the longest of the remaining five RTs was only equal to the standard error of the difference (approximately 1.6 msec. in both cases).

TABLE 2

MEAN REACTION TIMES CLASSIFIED BY FOREPERIOD DURATION

Flash Luminance in Ft-L	S	Foreperiod Duration		
		1.5 Sec.	2.0 Sec.	2.5 Sec.
3000	EF	163.4	155.9	155.4
	MH	165.1	150.9	147.4
	DR	168.7	158.8	156.0
	Mean	165.7	155.2	152.9
30	EF	176.1	169.8	169.0
	MH	176.3	162.6	158.7
	DR	179.2	168.6	165.4
	Mean	177.2	167.0	164.4
0.30	EF	202.7	196.1	195.0
	MH	202.2	188.8	181.8
	DR	212.6	201.1	195.8
	Mean	205.8	195.3	190.9

Note.—Each individual mean is based on 132 measures. The σ_m of the individual means range between 0.87 and 1.68 msec.

The data in Table 1 show clearly that luminance does affect RT. The RT to the 3000-ft-L light was, on the average, 12 msec. faster than to the 30-ft-L light, and 39 msec. faster than to the 0.30-ft-L light. As the standard error for each *S*'s combined mean for a given luminance was less than 1 msec., these differences are highly significant for each *S*.

The effect of foreperiod on RT is shown in Table 2. In all cases, length of RT is inversely related to foreperiod duration. In fact, for the three *S*s together, the difference in RT between the longest and shortest foreperiods was slightly greater than the difference between the 3000- and the 30-ft-L lights. For each *S*, at each of the three flash luminances, the difference between the 1.5-sec. and the 2.0-sec. foreperiod means was significant well beyond the .001 level. One *S* exhibited no difference

between the 2.0- and the 2.5-sec. foreperiod means. For each of the other two *S*s, the means at each luminance for the 2.0-sec. foreperiod were significantly longer than those for the 2.5-sec. foreperiod, the six *t*'s ranging from 2.14 to 4.14. The foreperiod effect found here is similar to that reported recently by Karlin (1959) and Klemmer (1956), both of whom showed that, over a given range of foreperiods, RT decreases with foreperiod length.

It is interesting to note that the effect of foreperiod on RT is independent of luminance level. Thus, for the three *S*s together, the differences in RT between the shortest and medium foreperiods are 10.5, 10.2, and 10.3 msec., respectively, for the high, middle and low levels. The corresponding differences between the medium and long foreperiods are 2.3, 2.6, and 4.4 msec.

Experiment II

Method

Subjects.—Fifteen undergraduates and one instructor, all naive in respect to the problem, served as *S*s. The *S*s were assigned at random to two groups of eight each, one a block and one a mixed group. Each *S* served in eight sessions.

Procedure: block group.—In this experiment, only the 50- and 250-msec.-flash durations of the 3000-ft-L light were used, as in the study of Baumgardt and Segal (1942). The two flash durations were presented in blocks of 18 trials each. Three foreperiods were used as in Exp. I. A daily session consisted of four such blocks in ABBA or BAAB sequence, these sequences being alternated on succeeding sessions in order to cancel any daily work curve effects. After 5 min. of dark adaptation, eight practice trials were given before the recorded trials of the first block were run. A 1-min. rest period separated successive blocks. The longest and the shortest RT of each block were discarded before block means were computed. The means for each flash duration for all eight sessions for a given *S* were thus based on 256 reactions.

Procedure: mixed group.—A session consisted of four sets of 18 trials each. In each set, the 50- and the 250-msec. flashes were presented nine times each in random order, three times at each foreperiod. Otherwise, the procedure was identical with that of the block group. As each set contained only nine reactions to a given flash duration, the longest and shortest RTs were not discarded in computing the means. Each duration mean for an *S* for all eight sessions was therefore based on 288 RTs.

Results

For the mixed group, flash duration had virtually no effect on RT. The data in Table 3 show almost identical RTs for the 50- and 250-msec. durations. Moreover, no individual reacted significantly faster to either stimulus. The data are in complete agreement with those of Exp. I.

Two of the eight *S*s of the block group reacted significantly faster to the 50-msec. flash and five others reacted faster but not significantly so. Only one of the eight showed a reversal of this trend. A test of the significance of the eight differences yields a *t* of 1.92 which by a one-tailed test is significant at the .05 level for 7 *df*. By a sign test, 7 of 8 differences in the same expected direction is significant at the .04

TABLE 3

Mean Reaction Times under Block and Mixed Presentations

Block Group				Mixed Group			
S	50 msec.	250 msec.	Diff.	*S*	50 msec.	250 msec.	Diff.
1	165.1	170.4	5.3*	1	166.4	167.6	1.2
2	165.3	168.9	3.6*	2	162.1	162.3	0.2
3	157.1	158.9	1.8	3	153.8	153.8	0.0
4	159.2	160.8	1.6	4	164.8	164.7	−0.1
5	169.5	170.4	0.9	5	174.5	174.3	−0.2
6	147.3	148.1	0.8	6	174.2	173.7	−0.5
7	185.5	186.0	0.5	7	176.8	176.1	−0.7
8	174.0	171.8	−2.2	8	186.8	185.1	−1.7
Mean	165.4	166.9	1.5	Mean	169.9	169.7	−0.2

* $P < .01$.

level. Although the difference of 1.5 msec. in RT in favor of the 50-msec. flash is very small, it must be concluded that under block conditions of presentation, the briefer flash does yield faster RTs.

Discussion

The data of these experiments show that RT to distinctly suprathreshold light flashes is unrelated to flash durations between 10 and 500 msec. when durations are presented in random order. Results showing faster RT to 50-msec. flashes than to longer ones may, therefore, be attributable to attitudinal factors that can develop during blocks of trials. The fact that *Ss* in our block procedure showed smaller differences in RT to the two flash durations than those reported in previous studies may be due to the procedure we used. Our procedure differed from that of Wells and that of Chernikoff et al. in that we used varying foreperiods and a constant repetition rate whereas they used constant 2-sec. foreperiods and a varying repetition rate. Baumgardt and Segal specify neither rate of stimulus presentation nor foreperiod duration. It is conceivable that attitudinal factors affecting RT are sensitive to procedural differences of this sort.

Our data give no support to the thesis of Baumgardt and Segal that short flashes are reacted to faster because they appear brighter.

Early in our experiments it became apparent that our method of stimulus presentation yielded a brightness-duration relation very different from that of Broca and Sulzer. According to the latter, the 50-msec. flash should have appeared brighter than the longer flashes, especially when the stimulus luminance was high. Instead, we found that apparent brightness increased with stimulus duration up to the 500-msec. flash for all three luminance levels. As this observation was so at variance with the Broca-Sulzer phenomenon, the following brief experiment was undertaken to measure apparent brightness as a function of flash duration by the method of single stimuli. It should be pointed out that this method is quite different from the one that has been used in all demonstrations of the Broca-Sulzer effect. It has previously been measured by matching the brief flash with a flash of much longer duration, exposed to an adjoining retinal area. In our experiment, there was no comparison stimulus. The flash was presented alone, as in the RT experiments, against a feebly illuminated and constant background.

TABLE 4

Apparent Brightness Ratings of 3000-Ft-L Flash at Each Duration

S	Flash Duration in Milliseconds					
	10	25	50	100	250	500
1	4.4	5.3	6.5	7.0	8.1	8.8
2	1.7	1.8	3.3	4.4	7.1	8.8
3	2.8	3.7	5.3	6.6	8.5	9.0
Mean	2.9	3.6	5.1	6.0	7.9	8.9

Three undergraduate *Ss* were asked to rate the apparent brightness of the 3000-ft-L light on a 9-point scale, 9 representing very bright, 1 very dim, and 5 medium. The six flash durations at this intensity were presented in random order 24 times each. Although all *Ss* were aware that duration was a stimulus variable, none suspected that intensity was held constant throughout. Instead, they reported that the variation in intensity was far more apparent than the variation in duration, and they experienced little difficulty in assigning the ratings.

The mean ratings by each *S* appear in Table 4. As the intensity of the stimulus was at a level that should yield a pronounced Broca-Sulzer effect, it is obvious that this application of the method of single stimuli yielded results altogether at variance with the matching procedures that have been employed by previous investigators of this effect. It is very possible that the Broca-Sulzer effect occurs only when a test flash of short duration is presented adjacent to or surrounded by a patch of light of several seconds duration, and whose onset precedes that of the test flash by a number of seconds. It may, therefore, depend on interaction effects originating in other retinal areas instead of being a sole function of flash duration.

Our experiments have shown that the continuation of a light flash beyond the

first 10 msec. (over the range of stimulus levels studied here) does affect phenomenal brightness but not RT. The reaction latency was apparently determined within the first 10 msec. of stimulation. Whether any other measure of response strength (such as force in the depression of the key) is correlated with phenomenal intensity cannot of course be answered on the basis of our data. In addition, it should be pointed out that our data give no answer to the question with which we began our investigations, namely, whether reaction time can be used as a measure of the Broca-Sulzer phenomenon. This phenomenon was not demonstrated under our conditions of stimulus presentation.

Summary

Previous studies of the effect of stimulus duration on RT have shown that RT increases as stimulus duration increases beyond approximately 50 msec. In Exp. I, in which six durations ranging from 10 to 500 msec. were presented in random order, RT did not vary with stimulus duration. RT was, however, found to be a decreasing function of luminance and of foreperiod duration. In Exp. II, in which 50- and 250-msec. flashes were presented in blocks of 18 trials each, a small but significant difference in favor of the briefer flash was found.

Our results suggest that attempts to explain the earlier findings on the basis of the Broca-Sulzer phenomenon are not justified. Instead, attitudinal factors that could develop when a given stimulus duration is presented repeatedly, as in the previous studies, are more likely to be responsible for the RT differences that have been reported.

An unsuccessful attempt to verify the Broca-Sulzer effect by the method of single stimuli indicates that this effect may require asynchronous stimulation of a neighboring retinal area. It was not observed in the absence of a comparison stimulus. Instead, phenomenal brightness increased with flash duration.

Under the range of conditions investigated, it is stimulus rather than apparent intensity that determines RT. The phenomenal brightness differences which were correlated with stimulus duration had no effect on reaction latency.

References

Alpern, M. Metacontrast. *J. Opt. Soc. Amer.*, 1953, **43**, 648–657.

Baumgardt, E., & Segal, J. Facilitation et inhibition parametres de la fonction visuelle. *Annee psychol.*, 1942, **43**, 54–102.

Bills, M. A. The lag of visual sensation in its relation to wave lengths and intensity of light. *Psychol. Monogr.*, 1920, **28**(5, Whole No. 127).

Broca, A., & Sulzer, D. La sensation lumineuse en fonction de temps. *CR Acad. Sci., Paris*, 1902, **134**, 831–834; **137**, 944–946, 977–979, 1046–1049.

Chernikoff, R., & Brogden, W. J. The effect of response termination of the stimulus upon reaction time. *J. comp. physiol. Psychol.*, 1949, **42**, 357–364.

Chernikoff, R., Gregg, L. W., & Brogden, W. J. The effect of fixed duration stimulus magnitude upon reaction time to a response terminated stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 123–128.

Crawford, B. H. Visual adaptation in relation to brief conditioning stimuli. *Proc. Roy. Soc. Lond.*, 1947, **134**, 283–302.

Gregg, L. W., & Brogden, W. J. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. comp. physiol. Psychol.*, 1950, **43**, 329–337. (a)

Gregg, L. W., & Brogden, W. J. The relation between reaction time and the duration of the auditory stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 389–395. (b)

Jarl, V. C. Method of stimulus presentation as antecedent variable in reaction time experiments. *Acta psychol.*, 1957, **13**, 225–241.

Johnson, H. M. Reaction-time measurements. *Psychol. Bull.*, 1923, **20**, 562–589.

Karlin, L. Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, 1959, **58**, 185–191.

Klemmer, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

Stainton, W. H. The phenomenon of Broca and Sulzer in foveal vision. *J. Opt. Soc. Amer.*, 1928, **16**, 26–39.

Stigler, R. Chronophotische Studien über den Umgebungskontrast. *Pflüg. Arch. ges. Physiol.*, 1910, **134**, 365–435.

Wells, G. R. The influence of stimulus duration on reaction time. *Psychol. Monogr.*, 1913, **15**(5, Whole No. 66).

(Received February 1, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 200–212

THE ORGANIZATION OF COMPONENT RESPONSE ERROR EVENTS IN TWO-DIMENSIONAL VISUAL TRACKING [1]

JACK A. ADAMS AND CARL E. WEBBER [2]

University of Illinois

The principal purpose of this paper is to test the generality of the law that time-on-target (TOT) response events on a trial in a component task of a multidimensional continuous tracking task are statistically independent of TOT response events occurring concurrently in other component tasks. For consistent terminology, the total multidimensional tracking task will be defined as having two or more stimulus sources, with each stimulus source having a corresponding dimension of the control system for response. Each stimulus source and its dimension of the control system will be considered a component task, and the component response is made to it. The total response is the joint action of all component responses.

Ellson (1947), using a three-dimensional pursuit tracking task, was the first investigator to infer that TOT response events on a component task were statistically independent of concurrent TOT response events on other component tasks. By treating proportion of TOT for a component task on a trial as a probability of success measure, and assuming the probabilities of all component tasks to be independent, we would expect their multiplication to result in the probability of joint success for the total task. Thus, if we had a two-dimensional tracking task with component tasks x and y, according to the hypothesis we should have $[P_{x(o)}][P_{y(o)}] = P_{xy(p)} = P_{xy(o)}$, where $P_{x(o)}$ and $P_{y(o)}$ represent observed proportion of TOT measured for the component tasks, $P_{xy(p)}$ the predicted probability that x and y are in simultaneous alignment, and $P_{xy(o)}$ the observed proportion of time that S is simultaneously on target in x and y. In effect, this relationship says that success in the total response, $P_{xy(o)}$, can be predicted from success on the parts. As an example, suppose that $P_{x(o)}$ and $P_{y(o)}$ are .30 and .50, respectively. The predicted probability of simultaneous TOT, $P_{xy(p)}$, is (.30)(.50) or .15, and would be compared with the empirical value $P_{xy(o)}$. Gardner (1950) followed Ellson's lead and gave the independence law some generality when data from a two-dimensional compensatory tracking task were found in support of it. Later research by Bilodeau (1955, 1957) and Bilodeau and Bilodeau (1954) added further support to the independence hypothesis in terms of both task and practice conditions, and they saw the law as a likely means for quantitative understanding of part-whole relations where the effects on $P_{xy(o)}$ could be predicted when $P_{x(o)}$ and $P_{y(o)}$ are changed by experimental treatments. Therefore, at present, we have no evidence to contradict the independence law and must accept it as a first quantitative principle of complex tracking behavior with no known restrictions on task, practice, or characteristics of the input signals. The purpose of this paper is to test the generality of this principle.

Table 1 illustrates independence and non-independence by showing the pattern of on-off target events of a particular scoring zone on a total tracking task with component tasks x and y, and is helpful in suggesting experimental approaches that might successfully challenge the independence law. One hundred hypothetical events are shown, with each event having the same time duration. The cell numbers not in parentheses illustrate frequencies fulfilling the conditions for statistical independence. Those in parentheses

[1] This research was supported by the United States Air Force under Contract No. AF 49(638)-371, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[2] Frank P. Serio and Wei Fong Ye were responsible for the development, calibration, and maintenance of the tracking apparatus, and Frank P. Serio kept the daily calibration of the analog-digital data system. Crayton C. Walker assisted in the data collection. William F. McDonald contributed to analysis of the data.

TABLE 1

ON-OFF TARGET EVENTS FOR A TRIAL IN A TWO-DIMENSIONAL TRACKING TASK
(An on-target event is represented by 1 and an off-target event by zero.)

		x = 1	x = 0	
y	1	20 (30)	30 (20)	50
	0	20 (10)	30 (40)	50
		40	60	N = 100

are an example of nonindependence. The probability of being on target in x, P_{x_1}, is .40, and the probability of being on target in y, P_{y_1}, is .50. The probability of success for the total response, $P_{x_1y_1}$, is .20 when events are independent and the product $(P_{x_1})(P_{y_1})$ equals it. But, when events are nonindependent, $P_{x_1y_1} = .30$ and the product term is an underestimation of it.

Tracking is fundamentally an error-nulling activity (Adams, 1961) and many man and machine variables i.e the total tracking system contribute to the pattern and magnitude of errors in relation to a particular TOT scoring zone. The independence law is based on a particular relationship among the frequency of error events shown in Table 1, and three variables suggest themselves as sensitive influences on tracking error for testing the generality of the independence law.

Characteristics of the input signal for each component task.—The percent TOT for each component task, P_{x_1} and P_{y_1}, is importantly influenced by the average difficulty of the input signal, and success of the total response, $P_{x_1y_1}$, is partly a function of the joint moment-to-moment pairings of cam segments which are relatively easy in terms of *S*'s proficiency level and width of the scoring zone. And the distribution of events in the other three cells in Table 1 are related to the frequency of segment pairings which are jointly difficult, or jointly easy and difficult. While one might conceive of many ways of generating input signals with a distribution of easy and difficult pairings of cam segments tending to promote the patterning of error events required for independence, it appears that one empirically proved way is to use unsystematic signals for each component task where the pairing of easy and difficult portions of each signal are "random" within the limits of finite cam mechanisms commonly used to generate inputs. Bilodeau (1955, 1957) and Bilodeau and Bilodeau (1954), who have amassed much of the evidence in support of the independence law, made use of the three-dimensional Stevens Pursuitmeter and the four-dimensional SAM Multidimensional Pursuit Task—both of which used unsystematic cam-based input signals for each component task. Nonindependence, however, has not been demonstrated empirically and it becomes necessary to conjecture how input signals might be designed to produce a failure of the multiplicative rule. Table 1 suggests a possibility. Notice that if there are disproportionate frequencies in the x_1y_1 cell and the x_0y_0 cell, nonindependence holds with $P_{x_1y_1(0)}$ underestimating $P_{x_1y_1(0)}$. One way of achieving this empirically in a two-dimensional tracking task is to have input signals which have greater likelihood for simultaneous on-target time (x_1y_1) or simultaneous off-target time (x_0y_0) than for either x_1y_0 or x_0y_1. A pair of signals which are at one time jointly easy for maximizing the accrual of x_1y_1 events, and at other times jointly difficult for maximizing the accrual of x_0y_0 events, but never one signal easy and the other difficult for x_0y_1 or x_1y_0 events, might generate an empirical demonstration of nonindependence. Rationally conceived input signals of this type may induce nonindependence but there is no guarantee because such variables as stage of practice or width of TOT scoring zone may have overriding influences and counter the expected effects.

Width of TOT scoring zone.—This variable defines the magnitude of tracking error that is classified as on-target and off-target events. The selection of the scoring zone is virtually always a semisubjective matter and only one zone has been used by previous investigators of this topic. Bahrick, Fitts, and Briggs (1957) have shown how the selection of a TOT scoring zone can influence the expression of functional relationships, and this may be true for the independence law also. A thoroughgoing test of the independence law should include an evaluation for a series of TOT scoring zones from small to large.

Practice.—As tracking proficiency changes with practice, the organization of error events in Table 1 can change. Practice, as a variable concomitantly manipulated with input signal and width of TOT scoring zone, may have effects on the independence law.

One other major consideration is the *behavioral* meaning of "independence" in this

tracking context. Ellson (1947, p. 12) has said that the independence law implies that *S* responds independently to each component task and does not correct errors simultaneously in two or more tracking dimensions with a single coordinated movement. This interpretation, however, must be regarded with caution. The x_1y_1 cell in Table 1 could have a high proportion of frequencies resulting from *S*'s ability to reduce two errors simultaneously with one movement, but the events still could be described by the multiplicative rule just as long as the remaining on-off target events in the other three cells stand in the proper relationship with each other and marginal totals for statistical independence. Nothing whatsoever is connoted for behavioral independence, i.e., that the corrective responses to eliminate error in multidimensional tracking are independent, like the tossing of coins. Probability mathematics, being a discipline of pure logic, has no empirical referents and says nothing of how events of the real world are related within an event class or between event classes in order that their relative frequencies are coordinate with statements of the probability calculus. Nagel (1947, p. 36) points out that the calculus of probability can be applied in many fields of research (e.g., the behavior of gases) even though the physical independence of events cannot be assumed to exist. A way of testing the intrinsic dependence among component response tracking error functions is to use cross-correlation (Merrill & Bennett, 1956) where two time series functions are repeatedly correlated over successively displaced intervals of time. Cross-correlation should reveal whether the error functions are related at the same or different moments in time even though their percent TOT scores conform to the independence law.

This experiment used a two-dimensional pursuit tracking task and manipulated characteristics of input signals, width of TOT scoring zone, and amount of practice to test the generality of the independence law. Cross-correlation as well as other intratrial analyses, are used to assess intrinsic dependencies between component response error functions.

METHOD

Tracking apparatus.—A two-dimentional pursuit tracking task was used. Each stimulus source was a rectangular aperture measuring 1 × 7 in. and mounted on a panel 28 in. from the seated *S*. Within the aperture, and on white background, was a black stimulus cursor driven by the cam input mechanism and a red response cursor activated by *S*'s movements of the control system. The maximum range of the black cursor was 5 in., but the red cursor traveled the full 7-in. scale. Each of the cursors appeared as a .10-in. wide reference line. The task of *S* was to strive for continual alignment of the red and black cursors. The rectangle of the left-hand stimulus source (*y*) was arrayed vertically and the right-hand rectangle (*x*) was horizontal. The control system was a two-dimensional stick similar to an aircraft control and was pivoted and spring-centered between *S*'s feet. For compatibility of control-display relations, the horizontal *x* source was connected to the left-right or *x* dimension of the control, and the vertical *y* source was connected to the forward-backward or *y* dimension of the control. Moreover, when the stick was moved left or right in the *x* dimension, the red cursor in the *x* source moved correspondingly to the left and right. And, when the stick was moved backward and forward in the *y* dimension, the red cursor in the *y* source moved up and down, respectively. The task was one of position tracking with a linear relationship in *x* and *y* between stick displacement and position of the red cursor. There was zero time delay between a step input to the stick and movement of the red cursor.

Performance measurement.—An analog-digital data system was used for recording the tracking error functions of each component task. Basically, an analog-digital data system rapidly and periodically samples each error voltage to a specified degree of accuracy, digitizes and encodes it, and permanently stores it on punched paper tape or magnetic tape. The digitized record is then submitted to a digital computer which, in conjunction with appropriate programs for the desired measures, will produce the performance measures. All of the measures traditionally used in tracking research, as well as many more, can be obtained from the output of a single unit of recording equipment. Our analog-digital data system required the distance between the red and black cursors of a component task to be represented by a linearly proportionate dc voltage. This analog voltage was fed to a digital voltmeter through a scanner which alternately sampled the error voltage level of each component task. The scale was 1-in. difference between

cursors equals 100 counts on the digital voltmeter, giving measurement accuracy of .01 in. on the display. Each reading was a three-digit number. The number, plus its algebraic sign, was coded as 1, 2, 4, 8 binary-coded decimal digits and serially punched on paper tape by a Teletype punch at the rate of 60 characters/sec. With each reading being four characters, the sampling rate was 15/sec. for the total system or 7.5/sec. for each component task. The raw data on the punched tape was appropriate for direct input to the digital computer Illiac, located at the University of Illinois.

Data reduction.—Digital recording of error voltages allows computation of TOT scores for any number of scoring zones desired, and in this study 17 zones are reported. The readings were recorded alternately on the tape which meant that strict simultaneity of events was not represented in the tape record. However, simultaneity was necessary for the computation of simultaneous TOT in the two component tasks, and the approach was to derive it from the recorded data. The serially recorded error scores looked like this: $y_1x_1y_2x_2 \cdots y_ix_iy_{i+1}x_{i+1} \cdots x_ny_n$. The best estimate of the value simultaneous with x_i was taken as the mean of y_i and y_{i+1}. This mean value was used in the computation of simultaneous TOT and TOT in y. Raw values were used for TOT in x. The use of the mean for achieving simultaneity carries the assumptions of equal time spacing of events and linearity of the analog voltage between two successively recorded points in the y dimension. Since successive values of y were recorded every 2/15 sec., and considering the response characteristics of S to a relatively slow stimulus input, it is reasonable to postulate linearity between two response measures separated by 2/15 sec.

The two limits of a scoring zone are symmetrical about zero error and are described by the number of digital counts from zero reference to the defining band of the zone. Thus, if a scoring zone is designated as 100, it means that S was scored on target whenever the red cursor was within $\pm$ 1 in. of the black cursor because the scale was 1 in. equals 100 counts on the digital voltmeter. For the computation of percent TOT for an S on a component task on a trial, the computer program prescribed an iterative process to compute the proportion of time that each successive pair of values was outside each of the 17 scoring zones. These time proportions were cumulated for each scoring zone, divided by the number of pairs of values, subtracted from one and expressed as percent TOT. Computation of simultaneous TOT entailed an assessment with respect to each scoring zone of two pairs of values, one from each component task, that were simultaneous in time. The proportion of time that either one or both were outside each of the scoring zones was determined, cumulated for all successive pairs in a zone, divided by the number of pairs of values, subtracted from one and expressed as percent simultaneous TOT.

Experimental design.—Two groups of 15 male *S*s each were used. Group I was expected to conform to the independence law and had cam input signals that were unsystematic for each component task, as well as unsystematic between component tasks. Within the limits of the cam following arm whose action actuated a potentiometer to produce the input signal, points of maximum and minimum on the cam, as well as rates between these points, were randomly selected. Group NI, for which nonindependence was hypothesized, had cams designed for testing the hypothesis that a disproportionate number of x_1y_1 and x_0y_0 events, with a reduced opportunity for x_1y_0 and x_0y_1 events, will produce inapplicability of the independence law. Group NI had each of the two cams smoothly cut with half of their circumference equidistant from the center, which caused the black cursors to remain stationary during half of each cam rotation. This provided a constant input to the two black cursors and displaced each of them a fixed distance from the center of the aperture for the duration of the smooth period. The responding for S during this time required him to hold each of the red cursors steadily on the black ones, with the purpose of maximally cumulating simultaneous TOT (x_1y_1 events). However, the other halves of the cams were irregular and asynchronized in their irregularities, making their signals quite difficult to track. During this segment it was anticipated that S would find it difficult to align in either component task, with the result that substantial simultaneous time-off-target (x_0y_0 events) would be tabulated. There were no cam subsegments where it seemed apparent that S would easily be on-target in one component task and off-target in the other for accruing x_1y_0 and x_0y_1 events.

The cams were not started at the same point on the beginning of each trial because it was felt that the initial ballistic movement of a typical S should be randomized with respect to cam segments. Cam rotation speed was not an even divisor of trial length so

cams ordinarily started each trial at a different place. The cam rotation speed for group I was 1.88 rpm and 2.79 rpm for Group NI. The faster speed for Group NI made the irregular half of the cams more difficult, while not disturbing the smooth portion, and this accentuated the difficulty of the irregular segments and increased the likelihood of occurrence of x_0y_0 events. The cam speed for Group I gave a total task of intermediate difficulty and was established by preliminary experimentation.

Twenty trials a day were given on each of three different days. Trial duration was 90 sec. and intertrial rest was 30 sec. Performance was recorded only on Trials 1, 4, 8, 12, 16, and 20 of each day as a means of keeping down the data mass and accompanying computer costs while, at the same time, providing enough data points to chart performance trends. An *S*'s awareness of the selective recording of his performance was minimized by having all cue-producing elements of the data system placed out of earshot in another room. There was no evidence that *S*s were aware of the occasional recording.

Subjects.—The 30 male *S*s were university undergraduates. They were paid for their participation. They were randomly assigned to groups.

Results

TOT Analysis[3]

The intention was to uncover failures in application of the independence law, if this is possible at all. An index of inapplicability is $[P_{xy(o)} - P_{xy(p)}]$, the difference between the observed and predicted simultaneous TOT probabilities. This difference may be either greater or less than zero when the independence law does not hold.

Group I.—Mean $[P_{xy(o)} - P_{xy(p)}]$ for a particular scoring zone on a trial was computed by multiplying $P_{x(o)}$ and $P_{y(o)}$ for each *S*, subtracting this product from his $P_{xy(o)}$, and then taking the mean of this difference score over all *S*s. This difference measure was computed for each trial and each scoring zone, and all were quite small with no evident pattern. Of the 306 mean differences, 62% were zero, 36% were .01, and 2% were .02. All were positive and none was larger than .02. The independence law applies unequivocally for Group I.

Group NI.—The mean $[P_{xy(o)} - P_{xy(p)}]$ values for Group NI are much larger and more systematic than for Group I, and are plotted in Fig. 1. In contrast to the negligible differences for Group I, Fig. 1 shows that the specially designed input signals were effective in disturbing the independence law and that the deviation of $[P_{xy(o)} - P_{xy(p)}]$ from zero (*a*) decreases as a function of practice, and (*b*) decreases as width of the scoring zone increases. The single exception to the latter relationship is the smallest Zone 5 which falls at about the same level as Zone 20.

Analysis of Intratrial Error Patterns

Cross-correlation was used to check the possibility that tracking error events of the two component tasks of Group I have some degree of intrinsic dependency even though their TOT scores are described by the independence law. This statistical method generates a number of correlation coefficients which, when plotted, show the degree of linear relationship between present error events in one component task with those of the other component task at the same or a later time. A cross-correlation function was computed for each *S* of Group I on Trial 60. The final trial was used because, if these dependencies exist and are learned, they should be most prominent on the terminal practice trial. The portion of the Trial 60 record used was generated during the second revolution of the cams. This portion was chosen to

[3] Available on request are the 17 acquisition curves of Groups I and NI in terms of mean probability of simultaneous TOT.

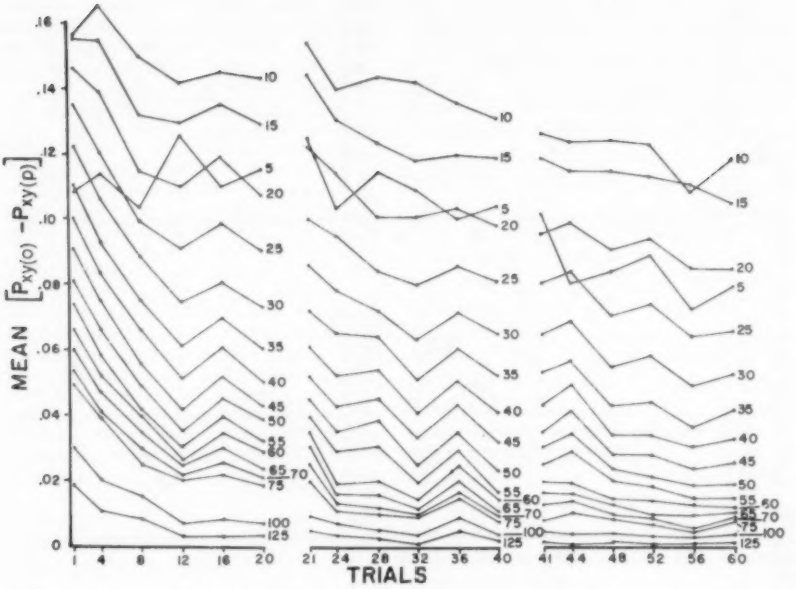

FIG. 1. For Group N1, the mean differences between predicted and observed probability of simultaneous time on target for 17 scoring zones.

avoid any atypical response effects that might arise from large errors occurring at the beginning of a trial when *S* first tried to align the cursors. The first r between the two time series had 240 values and was displaced by 1/15 sec., the time difference between any two successive readings on the data tape. The second, and all subsequent r's were successively displaced an additional 4/15 sec., until a maximum displacement of 16.07 sec. had been reached. The N for each successive r was reduced by two, until the final r had an N of 120. Sixty-one r's were computed for each *S* in this fashion and the plot of r's with respect to units of displaced time is the cross-correlation function.

Figure 2 shows the cross-correlation functions for six representative *S*s. The two reference lines for each function are the .05 confidence limits. Functions are presented for individual *S*s rather than a group mean function because the averaging of dissimilar individual functions effectively obliterated the important trends. The mean function was found to oscillate rather closely around zero but, as Fig. 2 shows, this is hardly the nature of individual curves which have a number of r's significant beyond the .05 level. Moreover, most of these significant r's are displaced in time, showing that the values of a component response error function at the present can be significantly related to those of the other error function at a considerably later time. Evidently it is possible for the component response error functions to have patterns of on-off target events for a TOT scoring zone describable by the

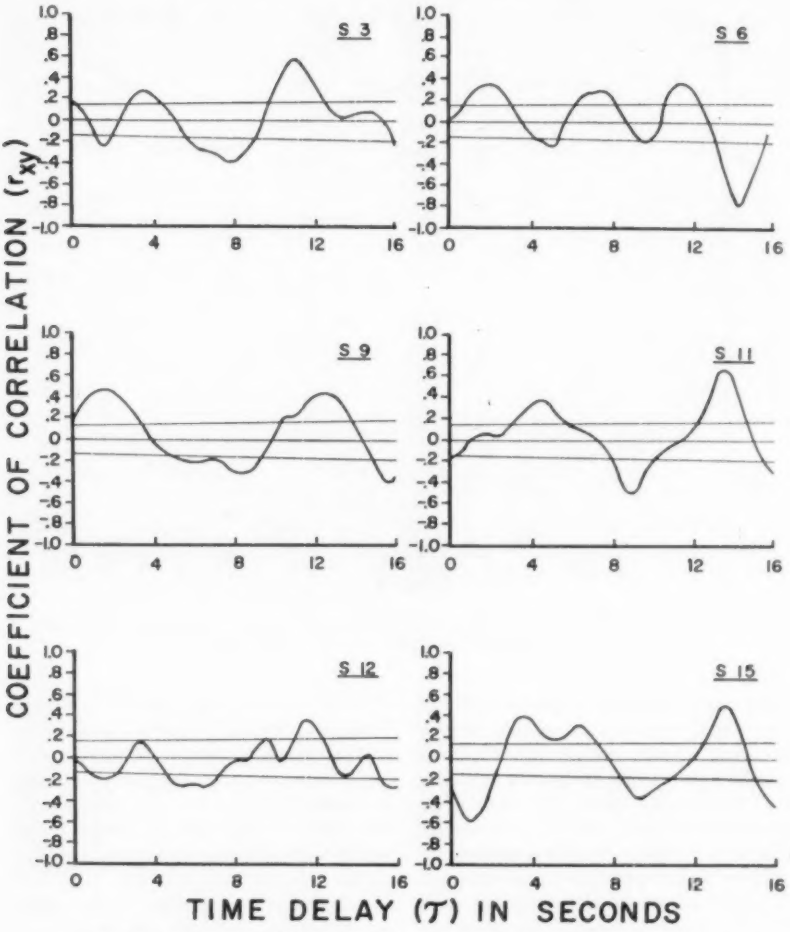

FIG. 2. Cross-correlation functions for six representative *S*s of Group I.

statistical rule for independent events while, at the same time, to have significant degrees of relatedness expressed by cross-correlation of the complete component response error functions.

In an attempt to better understand why the cross-correlation functions often revealed a significant relationship between tracking error events considerably displaced in time, the punched tape records for Trial 60 were printed out for visual inspection and analysis. A primary feature of these records which attracted our attention was that each component response error function was characterized by error peaks. Each error peak appeared as the rising of error values to a maximum, followed by a systematic decreasing of error magnitude to the vicinity of zero. The increase in error to initiate a peak can be a function of a number of task and response factors, but the reduction of error to *S*'s

criterion of excellence is almost solely a function of *S*'s visual attention to a stimulus source and his proper motor movements. While it is true that a fortuitous change of the input signal can cancel an existing error, *S* is the only systematic error-reducing entity in the system and, except for occasional chance factors, all error reduction can be ascribed to him. Thus, the presence of an error peak can be interpreted as the fundamental response act of sensing a tracking error and nullifying it with a correct motor movement. Examination of the Trial 60 records led to the conclusion that the coincidences of these error peaks, which are both negative and positive in relation to zero error, produced frequent excursions of the cross-correlation functions to significant values of *r*. As one component response error function is successively displaced in time with respect to the other and correlated, frequent occasions arise where the error peaks stand in the proper relationship to yield a significant positive or negative *r*.

The presence of error peaks raised the question of what their characteristics must be to fulfill the TOT independence law. Our approach was to devise a formal model about error peaks in two-dimensional tracking and to program the model for the Illiac computer and allow the machine to generate simulated tracking error records, just as our tracking task and analog-digital data system did in the laboratory. These simulated records were then subjected to the TOT program and conformity with the independence law was evaluated.[4] The model, which we call the Random Peak Model, had the basic assumption that many man and machine sources contribute to tracking error and create error peaks in each component response error function which have random rates of ascent to randomly selected amplitude maxima points, although rates of descent from the peak (*S*'s corrective act) are determined exactly by the amplitude of the peak. Values of one component response error function were always computed independently of those of the other function, but the computer output was a simulated tracking error tape where successive readings alternated between the two functions. The alternation gave the same format to the output tape as our laboratory analog-digital data system.

The model had the following mathematical properties:

1. The maximum point of a component response error function is defined as an error peak with an amplitude value E. The value E is randomly selected from a normal distribution whose scores have a mean of zero and ranged $\pm 3.7\sigma$. Plots on probability paper of the distributions of empirical error peaks for Trials 4 and 60 indicated only a moderate departure from normality.

2. The variable T is time taken by *S* to reduce E to zero. The function relating T to E is $T = aE^b$ where $b < 1$. The constants a and b were empirically estimated from our records. Contemporary studies relating response time to magnitude of movement suggest a function of this parabolic type (Brown & Slater-Hammel, 1949; Craig, 1949; Ellson, Hill, & Craig, 1949; Fitts, 1954; Searle & Taylor, 1948).

3. The quantity t was taken as the time for the error function to attain the value E, and was randomly selected from the parabolic function

[4] Our approach to building a model and assessing its consequences with a high-speed digital computer is of the sort ordinarily discussed under the Monte Carlo method. This procedure uses the computer to sample from one or more probability distribution functions, subject these values to one or more mathematical transformations, and generate a distribution of hypothetical response output values. An important application of the technique is for models whose output characteristics cannot be determined by formal, mathematical analysis.

relating E and T. To illustrate, the time t_i to attain E_i was taken equal to T_j found by substituting E_j in the parabolic function, where E_j was an error peak value different from E_i and was randomly chosen from the normal distribution of Es.

4. The error function is linear from zero error to E and from E to zero.

5. Successive readings for a component response error function are equally spaced in time.

The record generated by the model was equivalent to 10.7 min. of tracking and was submitted to the TOT program for analysis. The quantity $[P_{xy(o)} - P_{xy(p)}]$ was computed for each of 20 scoring zones. Seventeen of the differences were zero, and the remaining three were .01—indicating that the model can generate two-dimensional tracking records having the same close correspondence with the independence law as Group I.

Granting that an error peak is a basic tracking response act of sensing an error and executing a motor movement to eliminate it, the intratrial patterning of error peaks can show modes of attack and how behavior in a complex tracking task is organized. Moreover, it is of interest to see if the simulated peaks of the Random Peak Model approximate the patterning of empirical error peaks in addition to fitting the independence law. To evaluate these response patterning effects, all of the error peaks on Trial 4, Trial 60, and in the simulated tracking error data were identified. Trial 4 was included for information on trends from early to late in practice. The definition of an error peak was taken to be a point of error maximum that was followed by error decrease without reversal to within an error range of 0 ± 10, which was assumed to be closely related to the typical criterion of excellence for *S*s. The basis of this assumption was that the two cursors to be aligned were .10 in. wide (10 digital units) and the criterion of 0 ± 10 was a range within which separation of cursors was not apparent.

After the error peaks were identified, a probability analysis was performed. Four types of probability were computed:

(*a*) *Probability of Occurrence*—The relative frequency of an error peak in *x* or *y*. (*b*) *Probability of Repeating*—The conditional probability of an error peak in a component task being followed by an error peak in the same component task. (*c*) *Probability of Alternation*—The conditional probability that an error peak in one component task will be followed by an error peak in the other component task. (*d*) *Probability of Coincidence*—The conditional probability that the error peak of one component task will be coincident in time with an error in the other component task. Because recording of the error in the two component response error functions was sequential in time, the criterion of coincidence was error peaks occurring on immediately successive error readings.

Table 2 gives the results of the probability analysis. The Random Peak Model has reasonably good correspondence with the probabilities of Trial 60 but compares less favorably with Trial 4 by underestimating the Probability of Repeating and overestimating the Probability of Alternation for *x*. The model, by conforming to the independence law, fits the TOT data of Group I on all trials, but in terms of error peak patterns it corresponds only to advanced stages of practice. The behavioral tendency on Trial 4, which the model fails to respresent adequately, is a repetitive tendency, i.e.,

TABLE 2

PROBABILITY ANALYSIS OF ERROR PEAKS ON THE COMPONENT TASKS *x* AND *y* FOR GROUP I ON TRIALS 4 AND 60, AND FOR THE RANDOM PEAK MODEL

Measure	Trial 4		Trial 60		Random Peak Model	
	x	*y*	*x*	*y*	*x*	*y*
Probability Occurrence	.58	.42	.53	.47	.50	.50
Probability of Repeating	.48	.26	.31	.23	.25	.26
Probability of Alternation	.41	.59	.54	.61	.64	.63
Probability of Coincidence	.11	.15	.15	.17	.11	.10
No. of peaks for a component response error function	826	603	824	742	496	503
Total no. of peaks	1429		1566		999	

a kind of temporary response fixation on *x* before shifting to *y*. By Trial 60 however, the Probability of Repeating is diminished for *x* and the Probability of Alternation correspondingly increased, with the result that the patterning of error peaks on Trial 60 is about the same for *x* and *y*. On Trial 60, *S*s were more efficient as far as the total task is concerned because they improved their total response patterning by shifting to *y* more frequently after responding to *x*. The values for Probability of Coincidence show a very slight change with practice from Trial 4 to Trial 60, and they are of the same approximate magnitude as those of the model. The occurrence of coincident peaks could be due to the reduction of error through fortuitous cam action in one source at the same time that *S* is actively attending to the other source and reducing error in it. Coincident peaks also could be a function of precision visual-motor coordination where *S* simultaneously reduces two error quantities with a single integrated movement.

Table 3 presents additional characteristics of the error peaks on Trials 4 and 60, and elaborates some of the implications of the probability analysis in Table 2. Each *S* of Group I was given a score on number of error peaks, amplitude of error peaks, and intererror peak time interval, and the group means for these scores

TABLE 3

GROUP MEAN VALUES FROM ANALYSIS OF ERROR PEAKS OF COMPONENT TASKS *x* AND *y* OF GROUP I ON TRIALS 4 AND 60

Measure	Trial 4		Trial 60	
	x	*y*	*x*	*y*
Number of error peaks	55.1	40.2	54.9	49.3
Amplitude of error peaks (digital scores)	61.8	83.5	31.2	42.3
Intererror peak time interval (sec.)	1.65	2.36	1.63	1.85
$\Sigma\|e\|$ (summated tracking error in digital scores)	39648	49320	20653	24887

are presented in Table 3. The score for number of error peaks was, as implied, a counting of the number of peaks, but the other two scores for each *S* were means of their respective measures. The tendency on Trial 4 to repeat error corrections in *x* appears in Table 3 as a greater number of error peaks, a lower amplitude of error peaks, and a shorter time interval between peaks for *x* than for *y*. All of these differences between *x* and *y* on Trial 4 are significant at the .01 level. These characteristics of peaks would be expected to have association with an overall higher level of proficiency in *x*, and this is corroborated by a general tracking measure, summated tracking error $\Sigma|e|$ (same as the commonly used integrated error measure in the analog case except that discrete digitized scores are used). Table 2, however, shows that by trial 60, *S*s have reduced their tendency to fixate on *x* and repetitively correct, and now shift more frequently to *y* after having nullified an error peak in *x*. The increased attention to *y* results in a greater number of peaks for *y* (Table 3), although the increase in *y* from Trial 4 to Trial 60 falls just short of the statistical significance ($P = .07$). The greater number of peaks in *y* on Trial 60 necessitates a decrease in mean intererror peak time interval, and Table 3 shows a significant decrease from Trial 4 to Trial 60 ($.02 < P < .05$). Mean number of error peaks and mean intererror peak time interval in *x* are essentially unchanged with practice. The increase in number of peaks from Trial 4 to Trial 60 for *y* seems explainable partly in terms of learning effects revealed as reduced error peak amplitudes which require less time to correct. With less time needed for the correction of error peaks, more overall time becomes available and it is used to increase the number of corrective responses. Tables 2 and 3 show that whatever amount of time was freed as a result of the reduced time to respond was given over almost completely to additional corrections in *y*. By Trial 60 there were no significant mean differences between *x* and *y* in terms of number of error peaks and intererror peak time interval, although *x* had a mean peak amplitude that was smaller than *y* ($P = .02$), which is in the same direction as $\Sigma|e|$.

Discussion

Group I once again demonstrated the applicability of the independence law in multidimensional tracking when the input signal to each stimulus source is unsystematic, as well as freedom of the independence law from practice effects. Group I's data further extend the generality of the independence law by showing no relationship with width of scoring zone, an effect which has not been reported before. Group NI, however, proved conclusively that the independence law can be disturbed by specially designed input signals which redistribute the on-off target events and, when this is done, the applicability of the independence law becomes a function of scoring zone and practice.

Of particular interest is the cross-correlational analyses which revealed Group I's component response error functions as having intrinsic dependencies even though their data conformed to the independence law. This correlational finding argues against the hypothesis that the independence law implies independence of component response acts in multidimensional tracking (Ellson, 1947, p. 12). Additional evidence for dependency was found in our analysis of error peaks where substantial orderliness in responding was revealed by the level of conditional probability relationships among error peak occur-

rences. Considering all of these findings, we conclude that the independence law has limited generality for understanding the details of responding in complex tracking tasks.

A more fruitful approach to description of complex tracking behavior would seem to be in terms of error peak models, where an error peak is taken as a distinct response act of visually attending to a stimulus source and executing a motor movement to eliminate the observed error. By studying the probability structure of error peaks, and their time and amplitude characteristics, we can obtain insights into the organization of *S*'s behavior and how it relates to measures of overall tracking proficiency. Our response model was not an attempt at lawful description in the sense of linking independent and dependent variables. Rather we attempted to arrive at a succinct statistical model that would describe the general nature of our response data. Perhaps our effort should be viewed as a first step toward quantitative lawful description in that the Random Peak Model is one set of mathematical statements for dependent variables which the laws of multidimensional tracking must predict. If further research corroborates the usefulness of error peaks, then it will prove desirable to begin relating variations in the parameters of the model to independent task and procedural variables as a start toward expressing the laws of behavior in multidimensional tracking tasks.

Summary

An experiment was conducted to test the generality of the independence law which asserts that the time-on-target (TOT) response events of each component task of a multidimensional tracking task are statistically independent of one another. By this law the probability of TOT in the total task, defined as the simultaneous TOT for all component tasks, can be predicted by multiplying the probabilities of TOT in each of the separate component tasks.

A two-dimensional pursuit tracking task was used. Applicability of the independence law was tested as a function of characteristics of the input signals, width of the TOT scoring zone, and amount of practice. Group I (15 *Ss*) had unsystematic input signals and past research in this area suggested that its data should be describable by the independence law. Group NI (15 *Ss*) had input signals specially designed to induce inapplicability of the independence law.

The independence law fit Group I's TOT measures on all trials and all scoring zones. For Group NI, however, substantial discrepancies were found when the independence law was applied, and these deviations were a function of practice and width of the scoring zone. Time series cross-correlational analyses revealed significant dependencies among the component response error functions of Group I even though the probability mathematics for independent events fit the TOT scores. A further analysis of intratrial dependencies was made of error peaks, where an error peak was defined as the occurrence of an error followed by *S*'s visual sensing of it and making a motor movement to eliminate the observed error. The consequences of a Random Peak Model were assessed with a digital computer by the Monte Carlo method, and enough correspondences were found between the model, the independence law, and error peaks to suggest the fruitfulness of this approach for future work.

References

Adams, J. A. Human tracking behavior. *Psychol. Bull.*, 1961, **58,** 55–79.

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. Learning curves—facts or artifacts? *Psychol. Bull.*, 1957, **54,** 256–268.

Bilodeau, E. A. Variations in knowledge of component performance and its effects upon part-part and part-whole relations. *J. exp. Psychol.*, 1955, **50,** 215–224

Bilodeau, E. A. Patterns of internal consistency in multipart skilled performances. *Amer. J. Psychol.*, 1957, **70,** 550–559.

Bilodeau, E. A., & Bilodeau, I. McD. The contribution of component activities to the total psychomotor task. *J. exp. Psychol.*, 1954, **47,** 37–46.

Brown, J. S., & Slater-Hammel, A. T. Discrete movements in the horizontal plane as a function of their length and direction. *J. exp. Psychol.*, 1949, **39,** 84–95.

Craig, D. R. Effect of amplitude range on duration of responses to step function displacements. *USAF Air Materiel Command tech. Rep.*, 1949, No. 5913.

ELLSON, D. G. The independence of tracking in two and three dimensions with the G. E. Pedestal Sight. *USAAF Air Materiel Command memo. Rep.*, 1947, No. TSEAA-694-2G.

ELLSON, D. G., HILL, H., & CRAIG, D. R. The interaction of responses to step function stimuli: II. Equal opposed steps of varying amplitude. *USAF Air Materiel Command tech. Rep.*, 1949, No. 5911.

FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. exp. Psychol.*, 1954, **47**, 381–391.

GARDNER, J. F. Direction of pointer motion in relation to movement of flight controls. *USAF Air Materiel Command tech. Rep.*, 1950, No. 6016.

MERRILL, W. J., & BENNETT, C. A. The application of temporal correlation techniques in psychology. *J. appl. Psychol.*, 1956, **40**, 272–280.

NAGEL, E. Principles of the theory of probability. In, *International encyclopedia of unified science*. Vol. 1. Chicago: Univer. Chicago Press, 1947.

SEARLE, L. V., & TAYLOR, F. V. Studies of tracking behavior: I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.

(Received January 21, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 213–217

THE EFFECTS OF ANXIETY LEVEL AND SHOCK ON A PAIRED-ASSOCIATE VERBAL TASK[1]

LEE CHARLOTTE LEE

Fels Research Institute, Yellow Springs, Ohio

In recent years a number of experiments have attempted to verify Hull's basic assumption that habit (*H*) and drive (*D*) jointly determine response strength. Spence and Taylor have tested this assumption by treating anxiety, defined by the Taylor Manifest Anxiety Scale (MAS), as a measure of *D* (Taylor, 1953). The *Ss* scoring high on this scale are assumed to have high *D* (Hi-A *Ss*), and *Ss* receiving low scores are assumed to have low *D* (Lo-A *Ss*) in the experimental situation.

In the area of paired-associate verbal learning, Spence, Farber, and McFann (1956), Spence, Taylor, and Ketchel (1956), Taylor and Chapman (1956), Spence (1956) have found that Hi-A *Ss* perform significantly better than Lo-A *Ss* in tasks where the *H* of the correct response was stronger than the incorrect response. They also found Hi-A *Ss* to be inferior to the Lo-A *Ss* when the *H* of the incorrect response was stronger. Korchin and Levine's (1957) findings were in agreement in the latter task, but they were unable to verify the results in the former situation. Recent reports have indicated that psychological stress (Gordon & Berlyne, 1954; Lucas, 1952) and shock (Chiles, 1958) had the same effects as the MAS in the verbal performances of Hi-A and Lo-A *Ss*. However, Deese, Lazarus, and Keenan (1953) reported the use of intermittent shock significantly impaired performance of the Lo-A *Ss*, while shock had no effect on the Hi-A *Ss*.

Most of the experimental designs in the area of verbal learning used only normative data in assuming the relative habit strength of the correct and incorrect responses. These norms only loosely define the level of *H* on the particular pairs used.

Castaneda (1956; Castaneda & Palermo, 1955) used a different procedure in defining the relative *H* of S-R pairs in a psychomotor task. The *Hs* were established in *Ss* by giving them preliminary practice on particular sets of S-R pairs that were used in the learning task. The task was then modified by changing some of the pairs while others were left unchanged in transfer trials. This design is advantageous since it insures that all groups during the transfer period have the postulated pre-existing *H*.

The present study tested the $H \times D$ hypothesis using the MAS and shock as measures of *D* in a verbal learning situation in which the transfer list contained three types of S-R pairs: new S-R pairs, changed S-R pairs, and unchanged S-R pairs.

METHOD

Subjects.—The *Ss* were 60 men in elementary psychology classes at Ohio University. They were selected on the basis of scores they achieved on the MAS (Taylor, 1953). The distribution of scores ranged from 2 to 31 with a mean of 14. The Hi-A *Ss* had raw scores of 23 and above (upper 20%) and the Lo-A *Ss* had raw scores of 7

[1] This study is based on a thesis submitted to the Graduate School, Ohio University in partial fulfillment for the Master of Science degree. The author is indebted to Paul A. Games for his supervision.

TABLE 1

LIST OF S-R PAIRS FOR INITIAL AND TRANSFER TRIALS

Initial		Transfer	
Stimulus	Response	Stimulus	Response
[a]Absurd	Stupid	Distant	Yonder
[a]Pretty	Dainty	Tranquil	Composed
[a]Faulty	Unsound	Polite	Gracious
[a]Liquid	Solvent	Lazy	Inert
[a]Barren	Sterile	Modest	Bashful
Afraid	Timid	Brutal	Obscure
Frigid	Chilling	Unclear	Rustic
Former	Bygone	Speedy	Unkind
Winding	Twisted	Fickle	Hurried
Double	Duplex	Rural	Flighty
Brutal	Unkind		
Unclear	Obscure		
Speedy	Hurried		
Fickle	Flighty		
Rural	Rustic		

[a] Used also in transfer trials.

and below (lower 20%). The groups of 30 Hi-A and Lo-A *Ss* were subdivided randomly but equally into a Shock group and No-Shock group with resulting *Ns* of 15 in each of four groups. Three *Ss*, two Hi-A and one Lo-A, were discarded for failure to meet the established criterion of learning in the initial learning trials. One *S* in the Hi-A Shock group was discarded due to abnormal emotional reactions to shock in the transfer trials and refusal to continue in the experiment. One Lo-A *S* was discarded because of an eye infection.

Apparatus and materials.—Twenty pairs of adjectives (hereafter indicated as S-R pairs) from the Haagen (1949) list of common adjectives were used. Four lists of five pairs, equated for association value (closeness of association between S-R pairs), familiarity, and length, were constructed. The lists are presented in Table 1. The mean association value ranged from 2.2 to 2.6. A modified Lafayette memory drum was used to present the S-R pairs.

The shock apparatus consisted of a variac transformer connected to two electrodes made with silver dimes. The electrodes were fastened on *S*'s wrist approximately 1 in. apart by means of an elastic wrist band. For protection, the shock circuit went through two 6.3-v. transformers back to back and fuses of .5 amp. The shock duration of 1 sec. was controlled by a Hunter electric timer Model 115.

Procedure.—The *Ss* were seated at a table across from *E* with a board shielding *E*. The anticipation method was used in presenting the S-R pairs with both stimulus and response words exposed for 2.2 sec. The *Ss* were asked to name colors off the drum during the 11-sec. interval between each trial.

Each *S* received one presentation trial (*Ss* were asked not to respond on this trial) and four initial learning trials on the 15 S-R pairs. The order of the pairs was randomized in three different sequences with the restriction that no two words beginning or ending with the same sound might appear in succession. All *Ss* had to give at least seven correct responses in one of the four initial learning trials or they were discarded.[2]

After the four initial trials, the Shock and No-Shock *Ss* were given a 2-min. rest period. At the beginning of this rest period, electrodes were placed on shock *Ss*' wrist and the shock was gradually increased until *S* reported pain. The remainder of the 2-min. rest period was spent in naming colors that were pasted on a board beside the memory drum. The No-Shock *Ss* named colors throughout the 2-min. period without interruption.

The seven learning trials on the transfer list then followed, the first of which was a presentation trial. The instructions to the Shock and No-Shock groups were different; the Shock groups were told that they were going to be shocked regardless of their performance. Shock was not mentioned to the No-Shock groups. The procedure for all groups was similar to that used in the initial learning trials except that a new list of S-R pairs was used and the shock was administered to half of the *Ss*. The Shock *Ss* received a 1-sec. shock during the intertrial interval (in order to avoid association with specific responses). The shock was always administered approximately 4.4 sec. after the last S-R pair was removed so that 5.6 sec. of the intertrial interval elapsed between the end of the shock and the start of the next trial.

The transfer list consisted of 15 pairs of

[2] The length of the initial learning period was set at four trials as a result of two pilot studies to reduce the possibility of ceiling effects on the Unchanged pairs. If all *Ss* have five correct responses to the five Unchanged pairs during the second learning period, it would obviously be impossible for Hi-A *Ss* to be superior to the Lo-A *Ss* in regard to performance on the Unchanged pairs.

TABLE 2

MEAN NUMBER OF CORRECT RESPONSES IN TRANSFER TRIALS 2–7

Group	S-R Pair Categories					
	Neutral		Unchanged		Changed	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Lo-A No-Shock	19.14	3.85	19.80	4.41	7.67	4.16
Hi-A No-Shock	19.34	4.28	21.67	3.69	5.80	4.79
Lo-A Shock	16.94	6.09	18.40	4.91	3.26	2.50
Hi-A Shock	20.67	4.37	23.54	5.15	4.20	3.08

adjectives. Five of the pairs were completely new; these are called Neutral pairs. Five of the pairs were the same as in the initial learning trials; these are called Unchanged pairs. The remaining five pairs were made by re-pairing stimuli from the original lists with responses from the original list. These are called Changed pairs (see Table 1).

The sequence of presentation for the new list was also randomized with the same restrictions as mentioned above in the original learning, with the additional restriction that no more than two words in the same category (Changed, Unchanged, and Neutral) might appear in succession. The *Ss* were informed that some of the pairs were changed, some remained the same, and some were completely new.

All responses given were recorded as either correct or incorrect. A response was considered to be correct when *S* gave the correct response within the 2.2-sec. stimulus presentation interval. Record was also kept of competing responses. Competing responses were defined as any response word used incorrectly in the transfer trials that was on the response list in the initial learning trials. This included responses that were (*a*) incorrectly given to new stimuli of the Neutral pairs and (*b*) given to the wrong stimulus on the Changed or Unchanged pairs, although this response was still correct for another stimulus. (This included responses incorrectly retained on the Changed pairs.)

RESULTS

The total number of correct responses made on the initial learning trials was determined for each *S*. The means and variances of the Hi-A and Lo-A groups were not significantly different ($t = .119$, 58 *df*).

The total number of correct responses for each Pair Category were summated over the last six transfer trials. The mean number of correct responses of each of the four groups over the three different Pair Categories (Changed, Unchanged, and Neutral) is presented in Table 2. The Hi-A Shock group gave more correct responses when the pairs were Unchanged than the Lo-A No-Shock group. However, in the Changed pairs, the Hi-A Shock group gave fewer correct responses than the Lo-A No-Shock group. The Lo-A Shock group gave fewer correct responses than the other three groups in all the Pair Categories. A test for homogeneity of variance within the groups for transfer trials did not yield any significant differences ($\chi^2 = 3.87$, 56 *df*).

The summary of an analysis of variance (Lindquist, 1956, p. 281) on the transfer trials is presented in Table 3. The main effects of Anxiety, Shock, and the Anxiety × Shock interaction were not significant. However, the main effect of Pair Categories was significant. The interactions of Pair Categories and Drives (in terms of Shock or Anxiety) were significant. The triple interaction of Pair Categories, Anxiety, and Shock was not significant.

Six *t* tests for independent groups

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT RESPONSES IN TRANSFER TRIALS 2–7

Source	*df*	*MS*	*F*
Between-*Ss*			
Shock (S)	1	51.20	.03
Anxiety (A)	1	125.00	.08
S × A	1	115.20	.07
Error (b)	56	1,575.93	—
Within-*Ss*			
Pair categories (PC)	2	4,372.25	369.90***
PC × S	2	43.69	3.70*
PC × A	2	59.99	5.08**
PC × S × A	2	.54	.05
Error (w)	112	11.82	—

* $P = .05$.
** $P = .01$.
*** $P = .001$.

were computed, three on the three Pair Categories of Hi-A vs. Lo-A and three on the Pair Categories of Shock vs. No-Shock. The results yielded only two significant *t*'s of 2.88 and 3.00 with 58 *df* for the Hi-A and Lo-A groups on the Unchanged pairs, and Shock and No-Shock groups on the Changed pairs, respectively.

The competing responses given by *Ss* were summated for the four groups, and a 2 × 2 factorial analysis was run. No significant differences were found. (This may be due to the low frequency of such responses. The *Ss* tended to remain silent instead of guessing despite *E*'s instructions that guessing would not be penalized.)

DISCUSSION

The initial learning trials were used in an attempt to establish *H* strength for each *S* on specific S-R pairs. rather than to rely on normative data. The only distinction between populations on these trials was that of *D* level based on MAS scores (Hi-A vs. Lo-A). The conditions of the initial learning trials were very similar to the learning situations in other experiments using high association connections between S-R pairs except possibly in intralist similarity. (*E* here did not formally attempt to control intralist similarity; however, pairs were discarded during two pilot studies whenever there was evidence that intralist competition existed.) Although there was a slight superiority in performance of the Hi-A *Ss* during this learning period, the difference was not significant.

The analysis of the mean number of correct responses in the transfer trials generally supports the predictions dictated by the $H \times D$ hypothesis. The interactions between the Pair Categories and the two *D*s, Anxiety and Shock, were significant, indicating that the effects of *D*, either in terms of Anxiety or Shock, on performance depend on the type of Pair Categories involved. In other words, the Hi-*D* *Ss* performed better than the Lo-*D* *Ss* on the Unchanged pairs when the correct *H* was dominant, but they performed poorer on the Changed pairs when the incorrect *H* was dominant. The Hi-*D* *Ss* also performed better than Lo-*D* *Ss* on the Neutral pairs when the correct *H* may be presumed to be dominant, although the difference between the two groups was not as pronounced as in the Unchanged pairs. This is in accord with the author's assumption of the limitations of a normative definition of habit strength since the initial learning trials increased or established the S-R habits for the Unchanged pairs, while the S-R habit for the Neutral pairs depended on the normative generalities only.

Although interactions of *D* and Pair Categories were significant, the separate analysis of the six possible *t*'s found only two to be significant. The Hi-A *Ss* performed significantly better than the Lo-A *Ss* on the Unchanged pairs ($t = 2.88$, $df = 58$, $P < .01$), and the Shock group performed significantly poorer than the No-Shock group on the Changed pairs ($t = 3.00$, $df = 58$, $P < .01$). It should be noted that such *t* tests involve a between-*Ss* error term, whereas the

interaction tests involve a within-*Ss* error term.

The nonsignificant triple interaction of Anxiety, Shock, and Pair Categories may be taken as support for the contention that anxiety as measured by the MAS can be considered chronic, i.e., present in experimental situations regardless of whether noxious stimuli are used. The Anxiety × Pair Category interaction under the Shock condition did not differ significantly from that found under the No-Shock condition. This would be expected if the MAS predictions were specific to noxious stimulus situations. It is concluded that the present results give support to the $H \times D$ hypothesis.

Summary

The present study was concerned with the effects of MAS Anxiety and Shock on the transfer learning of Changed, Unchanged, and Neutral S-R pairs. It tested the $H \times D$ hypothesis with respect to D as measured by the MAS, and D as defined by shock administered during the intertrial interval.

Sixty male students were selected from the upper and lower 20% of a distribution of MAS scores. After all *Ss* learned a list of 15 adjective pairs, half of the Hi-A and half of the Lo-A *Ss* learned a second mixed list without shock, and the other half learned the mixed list with shock between trials. The mixed list of 15 pairs was made from 5 Unchanged pairs (same S and R as initially learned), 5 Neutral pairs (new Ss and Rs), and 5 Changed pairs (initial list Ss and Rs but different pairings).

The results with the second mixed list indicated that an increase in D (in terms of Anxiety or Shock) significantly facilitated performance when the dominant H was correct. When the dominant H was incorrect, an increase in D significantly impaired performance. The results also gave evidence that Anxiety as measured by the MAS should be considered chronic, i.e., present in experimental situations regardless of shock or no-shock.

References

Castaneda, A. Effects of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52,** 9–12.

Castaneda, A., & Palermo, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50,** 175–179.

Chiles, W. D. Effects of shock-induced stress on verbal performance. *J. exp. Psychol.*, 1958, **56,** 159.

Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46,** 55–60.

Gordon, W. M., & Berlyne, D. E. Drive-level and flexibility in paired-associate nonsense-syllable learning. *Quart. J. exp. Psychol.*, 1954, **6,** 181–185.

Haagen, C. H. Synonymity, vividness, familiarity and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27,** 453–463.

Korchin, S. J., & Levine, S. Anxiety and verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54,** 234.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

Lucas, J. D. The interactive effects of anxiety, failure, and interserial duplication. *Amer. J. Psychol.*, 1952, **65,** 59–66.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W., Farber, I. E., & McFann, H. H. The relation of anxiety (drive) level to performance in competitional and non-competitional paired associates learning. *J. exp. Psychol.*, 1956, **52,** 296–305.

Spence, K. W., Taylor, J. A., & Ketchel, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp, Psychol.*, 1956, **52,** 306–310.

Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.

Taylor, J. A., & Chapman, J. P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68,** 671.

(Received January 27, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 218–221

THE PROBABILITY OF DETECTING SMALL NUMBERS OF DOTS

FRANK RESTLE, JAMES RAE,[1] AND CHARLES KIESLER
Michigan State University

The question of span-of-apprehension asks how many things the mind can apprehend in a single instant. It appears that one can easily see two or three dots, but not 15, in a single glance. The accuracy of judging the number of dots decreases the more dots are present (Hunter & Sigler, 1940). Schlosberg (1948) studied the Hunter-Sigler data and noticed that *O* would usually underestimate the number of dots by a fraction, producing a roughly bell-shaped distribution of judgments. Schlosberg then gave the Hunter-Sigler effect a probability interpretation. Suppose that n dots are presented and there is a probability p that any particular dot will be detected. Suppose further that the probability of detecting a dot is independent of what other dots are presented or detected. Then the probability of detecting all n dots is p^n, and in general the probability of detecting exactly r of the n dots is given by the binomial distribution

$$\binom{n}{r} p^r(1 - p)^{(n-r)} \qquad [1]$$

Schlosberg found that this theory agreed with the Hunter-Sigler data quite well. More direct test of the theory (Casperson & Schlosberg, 1950) was also favorable. On each trial, when n dots were presented, *O* was asked to state how many dots were present. Theoretically, *O*'s response corresponds to r and the binomial distribution gives the probability that the response will be r when n dots are presented. It was noticed, however, that on occasion *O* might give a theoretically impossible response by reporting more dots than were presented. To minimize such outcomes, *O* was instructed to report only dots he saw "for sure." This reduced the number of impossible reports by biasing *O*s toward reporting small numbers.

Casperson and Schlosberg's test of the probability hypothesis suffers from several conceptual difficulties. First, their theory takes no account of "false alarms," responses to dots which are not actually present. They attempted to minimize reports of absent dots, but it is doubtful that any procedure can eliminate them entirely without a great sacrifice in detection (Tanner & Swets, 1954). Second, since they ask only for the number of dots seen they have no way of knowing which dots are seen, and hence can make no direct test of the hypothesis that the probability of detecting a particular dot is independent of what other dots were present. Third, the response obtained is a judgment of numerousness, subject to possible biases, variable errors, etc., which would make the judgment different from the number of dots actually detected.

A solution to these problems is to determine which dots are detected in a display. Then one gets information about false alarms, and it is possible to determine if the probability of detecting a particular dot depends on what other dots are present. Also, in this case, one deals directly with detection responses rather than with judgments. We report here two ex-

[1] Now at the University of Michigan.

periments in which the detection of particular dots is observed, and in which the probability hypothesis can be subjected to close scrutiny.

Experiment I

In a preliminary experiment two of the authors, FR and JR, alternated as *O* and *E* in detecting dots in a tachistoscope. Black dots were located at the top, bottom, left, and right sides of white cards, and the targets consisted of all 16 combinations of from zero to all four dots. These targets were viewed under high illumination at an exposure time of 0.5 msec., and *O* reported which dots he believed were present, identifying them by position. Each *O* had 32 trials on the 16 cards. About 70% of the dots shown were detected and there were about 14% false alarms. One of the dots was much harder to see than the others because of glare resulting from uneven illumination. The over-all relative frequencies of detecting a dot with 0, 1, 2, or 3 others were .58, .70, .72, and .74, respectively, suggesting a process of mutual reinforcements between dots. False-alarm rates did not increase concurrently, being .17, .12, .15, and .11, respectively, for the four conditions. With only two *O*s and with observations which probably were not independent, we found no unexceptionable way of deciding whether this drift of detection probabilities was a real phenomenon. It was decided to repeat the study with a larger number of *O*s and some improvements of technique. The third author carried out this second experiment.

Experiment II

Method

Observers.—The *O*s were 16 students from an introductory course in psychology. These students participate in experiments as a part of their regular class requirements.

Apparatus and materials.—A Bell and Howell Robomatic slide projector was used to present slides on a 12 × 12 in. rear projection screen. A high-speed shutter was attached to the projector. Sixteen 2 × 2 in. slides were prepared with all combinations of from zero to four dots at the positions top, right, bottom, and left. The dots were 35% gray on 25% gray background as measured on the Kodak Gray Scale. When projected, the dots, ½ in. in diameter, appeared at points about 5 in. from the fixation point in the center of the screen.

Procedure.—Each trial consisted of presentation of all 16 slides, once each, in a prearranged random order which was different for each trial and the same for all *O*s. Each *O* received 10 trials. The first four trials were considered training trials, with exposure times of 1.0 sec., 0.5 sec., 0.2 sec., and 0.1 sec. These were followed by six test trials with all exposures at 0.1 sec. The aperture of the shutter was fixed on the basis of preliminary studies to yield detection probabilities about halfway between chance and perfect performance, and remained constant throughout the experiment. The *O* was not held to a rigid time schedule, but exposures were in fact separated by about 5 sec. After each two trials (32 exposures) a rest period of about 90 sec. intervened while the slides were rearranged in their magazines. A verbal pre-exposure warning and a central fixation point were provided.

The *O*s were tested individually or in groups of two or three at a time. Each *O* was seated 8 ft. from the screen in a classroom desk-chair with cardboard dividers obscuring each *O*'s view of the others.

The *O* on each trial reported which dots he believed had been presented, using the code names 1, 2, 3, 4 for the positions top, right, bottom, and left. Responses were made on standard IBM test answer sheets, and each *O* was carefully supervised to ensure complete understanding of the procedure before test trials began.

The six test trials gave a total of 96 observations for each *O*.

Results

The four positions were detected about equally often in this experiment.

TABLE 1

Proportion of Times a Dot Was Detected When Present, as a Function of the Number of Other Dots Displayed: Exp. II

Position of Dot Detected	Number of Other Dots Displayed			
	0	1	2	3
Top	.73	.58	.62	.59
Right	.75	.78	.79	.80
Bottom	.66	.66	.75	.65
Left	.70	.83	.75	.81
Mean	.71	.71	.73	.71

Note.—In addition, correct reports of zero dots were made on .61 of the presentations of zero dots.

TABLE 2

PROPORTION OF CORRECT DETECTIONS CORRECTED FOR FALSE ALARM RATE AS A FUNCTION OF THE NUMBER OF OTHER DOTS PRESENT: EXP. II

Position of Dot Correctly Detected	Number of Other Dots Present			
	0	1	2	3
Top	.70	.52	.53	.54
Right	.66	.74	.71	.76
Bottom	.59	.53	.61	.50
Left	.57	.75	.62	.70
Mean	.63	.64	.62	.63

Table 1 shows the relative frequencies of detection of dots in the various positions, as a function of the number of other dots present. It will be noted that there is no tendency for the proportion of detections to increase (or decrease) with the number of other dots present.

Each *O*'s data were corrected for guessing and the resulting corrected proportions were averaged. The results are shown in Table 2. From this table it is seen that there is again no tendency for corrected relative frequency of detection to vary with the number of other dots presented. In view of the absence of a trend, statistical tests on the data were not performed.

DISCUSSION

We have approached the classical problem of span-of-apprehension as a problem in simultaneous detection of several isolated targets. From this point of view we have asked our *O*s not only how many, but which dots they saw. As a result we have two observations to report: (*a*) the fairly high frequency of false alarms, which would be expected from the fact that we have a detection experiment, and (*b*) the finding that the probability of detecting a certain dot is not much affected by what other dots are present, for 0—4 dots.

The occurrence of false alarms raises a serious problem in the ordinary technique of asking *O* how many dots he sees. It is quite possible for *O* to report dots which are not there, and more dots than are shown. This makes it impossible to accept Schlosberg's probability model as stated above. According to that theory, the theoretical probability of detecting more dots than are shown must be zero. However, the probability of *reporting* more dots than are shown is not zero.

Aside from these technical difficulties, however, the general intention of the Schlosberg theory fares rather well when applied to our results. We have no repeatable evidence that the probability of detecting one isolated dot changes with the number of other dots presented. This suggests that the several detections may be independent random events, the main point of Schlosberg's conjecture.

Suppose there is a certain probability p that a dot will be reported if present, and a certain false-alarm probability p' that it will be reported if absent. If there are a total of N positions or possible dots, of which n are presented, one can calculate the probability that k will be reported, as follows. The k dots reported are composed of r of the dots presented and $k - r$ of the absent dots which are falsely reported. Each of the r dots is reported with probability p, and there are n such dots. The probability of reporting r of n dots is

$$\binom{n}{r} p^r(1-p)^{n-r}$$

At the same time, to have k reports, the subject must report exactly $k - r$ of the absent dots. There are $N - n$ dots absent, and probability p' that an absent dot is reported. The probability of reporting exactly $k - r$ false alarms is therefore

$$\binom{N-n}{k-r} p'^{(k-r)}(1-p')^{(N-n-(k-r))}$$

A report of exactly k dots may be made up of any combination of r detections and $k - r$ false alarms. For any particular r we multiply the two above expressions to get the probability of k reports, and then we sum these values for all possible values of r, from 0 to k.

As the probability of a report of k dots when n are present of a total of N possible dots, we have[2]

$$P(k; n, N) = \sum_{r=0}^{k} \left[\binom{n}{r} p^r (1-p)^{n-r} \right] \times \binom{N-n}{k-r} p'^{k-r} (1-p')^{N-n-(k-r)} \quad [2]$$

This expression, while considerably bulkier than the simple binomial distribution used by Schlosberg, has the advantage of admitting the possibility of reports of more dots than are present. Unfortunately it cannot be used directly on the usual span-of-apprehension experiments, because one has no clear indication of N, the number of positions or possible dots. This in turn indicates a slight conceptual looseness in the idea of having dots "randomly spaced" without specifying the sample space over which the distribution of dots is to be defined.

This study, which derives from the study of span-of-apprehension, seems to be a natural extension of the psychophysics of detection (the absolute threshold) to the case of more than one target. Provided that the various detections and false alarms are independent events, our Eq. 2 is useful in the analysis of such data. It is a purely mathematical problem to determine other probabilities which may be of interest, such as the probability that the correct number of dots is reported (that $k = n$), the mean and variance of the number of dots reported for a fixed number presented (n), and so forth. For example, the probability that at least one dot is reported, as a function of n, the number shown, is given by

$$1 - (1-p)^n (1-p')^{N-n} \quad [3]$$

One can imagine an application in, say, radarscope watching, in which the observer may be presented with zero, one, or more targets randomly arranged, and one asks the probability of his reporting at least one target as a function of the number of targets.

Extension of the results to the case of dots which are not independently detected, but which form patterns or which reinforce or inhibit one another, will no doubt introduce serious difficulties. The results of Exp. I suggest that *O* may see combinations of more than one dot somewhat more easily than he sees an individual dot. In other cases one might find competition between the various dots. It would be reasonable to seek such conditions by using more than four dots, by placing the dots close together, and by using suitable configurations of dots. The main results of Exp. II provide a base line from which interactions can be measured, and seem to open a new and promising approach to these deeper questions.

[2] This expression is the convolution of two binomial distributions. Let $a_1, a_2, \cdots, a_n$ be one probability distribution and $b_1, b_2, \cdots, b_n$ be another. Then $a_1 b_n + a_2 b_{n-1} + \cdots + a_n b_1$ is the convolution of the two (Feller, 1957).

Summary

In two experiments, *O* attempted to detect isolated dots under low illuminations and with short exposures. From zero to four dots at the positions top, left, bottom, and right were presented, and the locations of the dots were reported by *O*. False alarms were observed, and in the main experiment it was found that the probability of detecting any particular dot was not noticeably affected by the number of other dots presented. Schlosberg's "probability" theory of span-of-apprehension was reformulated to incorporate false alarms, and the problem was described as the detection of multiple targets.

References

Casperson, R. C., & Schlosberg, H. Monocular and binocular intensity thresholds for fields containing 1–7 dots. *J. exp. Psychol.*, 1950, **40**, 81–92.

Feller, W. *An introduction to probability theory and its applications.* Vol. I. New York: Wiley, 1957.

Hunter, W. S., & Sigler, M. The span of visual discrimination as a function of time and intensity of stimulation. *J. exp. Psychol.*, 1940, **26**, 160–179.

Schlosberg, H. A probability formulation of the Hunter-Sigler effect. *J. exp. Psychol.*, 1948, **38**, 155–167.

Tanner, W. P., & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Received January 27, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 222–227

SUBJECTIVE SIMILARITY IN TWO PERCEPTUAL CONTINUA [1]

GÖSTA EKMAN, GUNNAR GOUDE, AND YVONNE WAERN

University of Stockholm

Similarity as perceived by an *O* in psychological experiments or by anybody in daily life is a frequent attribute of psychological descriptions and has also often been used as an explanatory concept. In most of these cases there has been a certain confusion of objective and subjective similarity. A methodological analysis of the concept of similarity has been presented by Noble (1957). He concludes that similarity as a relation, is itself intrinsically nonmeasurable.

The present study, on the other hand, is concerned with subjective similarity, as it is naively perceived, and with the problem how this experience arises.

In a previous study this problem was investigated by Eisler and Ekman (1959) for the dimension of pitch. A ratio scale of pitch at constant loudness was first obtained. In another experimental series similarity estimates were obtained for all combinations of seven frequencies (ranging from 200 to 13,000 cps) of equal loudness. It was shown that subjective similarity (s_{ij}) between two tones was a simple function of the scale values (R_i, R_j) of the tones in the pitch continuum,

$$s_{ij} = \frac{R_i}{(R_i + R_j)/2} \quad (R_i \leq R_j) \quad [1]$$

This result means that subjective similarity in the dimension of pitch is equal to the ratio between the lowest scale value and the mean of the scale values of the two tones included in the comparison.

The study of pitch similarity was carried out as the first in a research project on mechanisms of similarity in unidimensional continua. The purpose of the present study is to investigate these mechanisms in two other perceptual continua: subjective darkness and subjective magnitude. The hypothesis of this study is that subjective similarity also in these continua may be adequately described by Equation 1.

EXPERIMENT I

This experiment was set up with the purpose of investigating similarity between percepts varying presumably only with regard to darkness.

Method

The stimulus material consisted of seven square cards, 55 × 55 mm. The cards were gray with varying degrees of darkness. In Hesselgren's color system (ranging from 0 = "white" to 24 = "black") the cards have the gray values 2, 4, 6, 8, 10, 12, and 14. (The light absorption of these cards is 35, 49, 58, 66, 72, 77, and 81%.) The cards were presented two at a time in an illuminated window 105 × 105 mm. The *O* was sitting in a dark room with his eyes in front of the window at a distance of 800 mm.

The experiment was carried out in two steps: (*a*) The *O* was instructed to estimate the degree of subjective similarity in each combination of two cards. The scale ranged from 0, denoting no similarity at all, to 10, denoting identity. The order of stimulus combinations was randomized, and right and left position was rotated. For each *O* four similarity estimates were obtained for each combination. (*b*) In order to obtain a ratio

[1] This study was supported by the Swedish State Institute of Psychological and Educational Research.

TABLE 1

SIMILARITY ESTIMATES IN EXP. I

	Stimulus Number					
	2	3	4	5	6	7
1	.91	.59	.46	.35	.22	.18
2		.82	.62	.51	.33	.27
3			.95	.80	.60	.56
4				.96	.77	.68
5					.89	.80
6						.98

scale of darkness, a method of ratio estimation (Ekman, 1958) was applied. The *O* was instructed to estimate the ratio of subjective darkness for each combination of two cards. The order of stimulus combinations was randomized and position rotated. Each *O* made four ratio estimates for each of the 21 combinations.

Five *O*s took part in the experiments, three men and two women. Their ages were 18, 22, 25, 20, and 46 yr. For two *O*s the ratio estimation preceded the estimation of similarity, and for three *O*s the order was reversed.

Results

The mean similarity estimates were divided by 10 and thus transformed to a scale ranging from 0 to 1, where 1 denotes identity. The estimates of similarity thus obtained are shown in Table 1. The mean ratio estimates are shown in Table 2.

The hypothesis of the present experiment was that the relation discovered in the case of pitch would apply also to the subjective dimension of darkness. In order to test this hypothesis, Equation 1 was written in the form

$$s_{ij} = \frac{2}{1 + R_j/R_i} \quad (R_i \leq R_j) \quad [2]$$

and theoretical values of similarity were obtained by applying this formula to the matrix of mean ratio estimates. These theoretical values are shown in Table 3.

Inspection of Tables 1 and 3 indicates a good agreement between experimental and theoretical values of similarity. This agreement is further illustrated by the squares in Fig. 1. A straight line has been fitted by the method of least squares to the data of this experiment. Its equation is

$$s_{\text{emp.}} = 0.005 + 0.98\, s_{\text{theor.}}$$

TABLE 2

RATIO ESTIMATES IN EXP. I

	Stimulus Number					
	2	3	4	5	6	7
1	.82	.37	.29	.21	.11	.08
2		.69	.46	.37	.21	.15
3			.93	.72	.49	.38
4				.93	.67	.57
5					.86	.72
6						.97

TABLE 3

THEORETICAL VALUES OF SIMILARITY IN EXP. I

	Stimulus Number					
	2	3	4	5	6	7
1	.90	.55	.44	.34	.20	.15
2		.82	.63	.53	.35	.26
3			.96	.83	.66	.56
4				.97	.80	.72
5					.92	.84
6						.98

TABLE 4

SIMILARITY ESTIMATES IN EXP. II

	Stimulus Number			
	2	3	4	5
1	.50	.25	.21	.09
2		.74	.46	.30
3			.82	.58
4				.87

The constants of this equation are rather close to 0 and 1, respectively. Thus the results support the hypothesis that subjective similarity in the dimension of darkness arises according to the same principle, indicated by Equation 1, as in the previously investigated case of pitch.

Experiment II

This experiment was designed to investigate similarity between percepts varying presumably only in the dimension of subjective visual area.

Method

Five circles with diameters of 10, 20, 30, 40, and 50 mm. were used as stimulus material. The circles were drawn with India ink on white cards, 150 × 210 mm., two on each card in all possible combinations so that there were 10 cards. The cards were presented to *O* one at a time. They were viewed from a distance of about 400 mm.

The experimental steps were the same as in Exp. I. (*a*) Similarity estimates were obtained on a scale ranging from 0 to 10. For each *O* four estimates were obtained. Upper-lower position was rotated and the order of stimulus combinations was randomized. (*b*) Ratio estimates were obtained as in the previous experiment. Each *O* made four estimates. Position was rotated and order of stimulus combinations was randomized in the first step.

Five *O*s took part in this experiment, one male and four female students of psychology, their ages being about 20 yr. The similarity estimation always preceded the ratio estimation in this experiment.

TABLE 5

Ratio Estimates in Exp. II

	Stimulus Number			
	2	3	4	5
1	.29	.18	.11	.08
2		.67	.31	.23
3			.75	.42
4				.80

TABLE 6

Theoretical Values of Similarity in Exp. II

	Stimulus Number			
	2	3	4	5
1	.57	.30	.21	.14
2		.80	.47	.38
3			.86	.59
4				.89

Results

The mean similarity estimates were divided by 10 and thus transformed to a scale ranging from 0 to 1 as in Exp. I. They are shown in Table 4. The mean ratio estimates are shown in Table 5.

Theoretical values of similarity according to the hypothesis were computed by applying Equation 2 to the matrix of mean ratio estimates. They are shown in Table 6.

The agreement between experimental and theoretical data is further illustrated by the triangles in Fig. 1. The best-fitting straight line for the data of Exp. II has the equation

$$s_{emp.} = 0.036 + 0.99\, s_{theor.}$$

According to this experiment, the hypothesis expressed by Equation 1 may be accepted at least as a good approximation.

Discussion

Since the results from the two experiments of this investigation show the same general trend as the results obtained in the previous experiment on pitch, the data from all thee experiments have been summarized in Fig. 1.

The plot exhibits a clear diagonal trend. This represents the hypothesis expressed in Equation 1, which means that subjective similarity between two percepts varying in a single perceptual dimension is equal to the ratio between

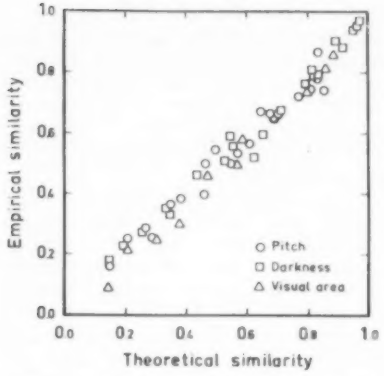

FIG. 1. Experimental estimates of similarity in three subjective continua, plotted against theoretical values according to Equation 1.

the lowest scale value and the arithmetic mean of the two scale values on this continuum, provided the measurements are expressed in a ratio scale. Since this relation has now been verified in three quite different perceptual continua, it seems possible that it is of a relatively general nature.

If Equation 1 is restated in the equivalent form $s_{ij} = 2R_i/(R_i + R_j)$, its psychological interpretation may appear even simpler, or at least more concrete. In this form it states that the similarity-generating process counts off that portion of R_j which corresponds to R_i and computes the ratio between the *common* subjective magnitude $(2R_i)$, and the *total* subjective magnitude $(R_i + R_j)$ present in the comparison.

Equation 1 is mathematically very simple. As was pointed out by Eisler and Ekman (1959), it may be characterized as an empirical equation without empirical constants. It will further be noted that the equation is entirely in *subjective* terms. This, in fact, is one of the rare instances where subjective mechanisms have been quantitatively investigated. Years ago, Guilford (1939) suggested research of this kind, which he called "psychodynamic" and which he illustrated by a color experiment. The present experiments indicate that intrasubjective relations may, sometimes, be quite simple even when the transformations from stimulation to subjective response are highly complicated.

The question may be raised, how similarity was measured in the present experiments. The method used in all these experiments may reasonably be considered a method of magnitude estimation, where the maximum similarity (identity) is arbitrarily assigned the scale value 10; dividing the estimates by 10 would accordingly yield scale values in terms of fractions of identity. Since the first experiment on pitch was undertaken, this question has been experimentally investigated. In a situation similar to Exp. II of the present study, Ekman and Waern (1959) measured similarity between circular surfaces by two methods. In one part of the experiments, similarity estimates were obtained by the same method of magnitude estimation as used in the present experiments. In another part of the experiments, *Os* had to compare all possible pairs of stimuli two at a time and to estimate, for each couple of pairs, the similarity of one pair in relation to the similarity of the other pair; this was a method of ratio estimation. The two scales thus obtained were approximately proportional to each other. Accordingly, similarity as estimated in these experiments may be considered measured on the ratio level.

Finally some possible implications of the similarity principle for scaling theory will be briefly discussed. As shown by Stevens and Galanter (1957) category scales deviate systematically from ratio scales for prothetic continua: when plotted against a ratio scale, category data exhibit a trend which is concave downwards. This means that a constant distance on the category scale corresponds to an increasing distance on the ratio scale. Stevens and Galanter explain this phenomenon by referring it to the fact that just noticeable differences in subjective units increase with position on the subjective continuum. The similarity principle may offer an

alternative explanation. Assume for the moment that O, working with a method of equal appearing intervals or any other category method, tries to fulfill the requirement that successive steps should be equal by making *successive pairs equally similar*,

$$s_{12} = s_{23} = \cdots = s_{(n-1)n}$$

From Equation 1 it is seen that in this case

$$R_1/R_2 = R_2/R_3 = \cdots = R_{(n-1)n}$$

i.e., equal steps or distances on the category scale will correspond to equal successive ratios on the ratio scale.

This principle is illustrated in Fig. 2 where a "category scale" (ordinate) has been constructed by letting successive pairs have a constant subjective similarity of 0.80, corresponding to a subjective ratio 2/3. This interpretation does not without further assumptions account for the bias introduced in category scales by various principles of stimulus grouping (Stevens & Galanter, 1957) but the alternative interpretation does not seem to be better off from this point of view. The problem cannot be solved without further experimentation.

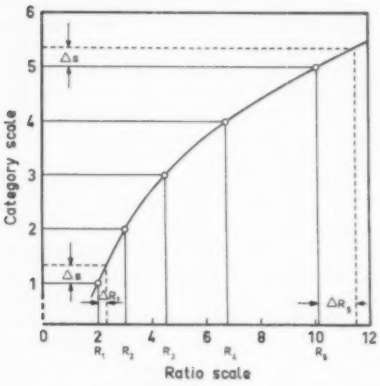

FIG. 2. Artificial category scale plotted against a ratio scale. The category scale was constructed on the assumption that "equal intervals" are obtained when successive pairs are made "equally similar." The diagram further illustrates that a just noticeable difference which is constant on the similarity continuum will increase with the subjective magnitude when measured on the ratio level.

The alternative interpretations discussed here may however be fundamentally identical. Let us finally assume that a just noticeable stimulus difference is constant on the similarity continuum. It can be shown from Equation 1 that the corresponding difference ΔR on the ratio level will increase with R, as is usually the case (cf. Ekman, 1959). This is also easily seen from Fig. 2.

In further experimentation on subjective similarity it may, accordingly, be of interest to consider also the concepts of "distance" and "just noticeable difference" as well as the relation between category scales and ratio scales. If the relation between category scales and ratio scales were completely described by the principles of similarity discussed here this relation would be a simple logarithmic function, which in general it is not. Similarity may, however, be one of several cues guiding O in his efforts to construct the interval scale required by E's instructions.

SUMMARY

The mechanisms of subjective similarity were investigated for the perceptual continua of darkness and subjective visual area. Similarity estimates were obtained and ratio scales of darkness and subjective area were constructed. It was shown that similarity between two percepts is equal to the ratio between the lowest scale value and the average of the scale values of the two percepts. This is the same relation as was found for the continuum of pitch investigated in a previous study by Eisler and Ekman. Some possible implications of the results for scaling theory are discussed. Experimental findings concerning differential sensitivity in subjective continua are tentatively interpreted in terms of the similarity principle.

REFERENCES

EISLER, H., & EKMAN, G. A mechanism of subjective similarity. *Acta psychol., Amst.*, 1959, **16**, 1–10.

EKMAN, G. Two generalized ratio scaling methods. *J. Psychol.*, 1958, **45**, 287–295.

EKMAN, G. Weber's law and related functions. *J. Psychol.*, 1959, **47**, 343–352.

EKMAN, G., & WAERN, Y. A second-order ratio scale. *Acta psychol., Amst.*, 1959, **16**, 78–80.

GUILFORD, J. P. A study in psychodynamics. *Psychometrika*, 1939, **4**, 1–23.

NOBLE, C. E. Psychology and the logic of similarity. *J. gen. Psychol.*, 1957, **57**, 23–43.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received January 30, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 228–235

STUDIES OF DISTRIBUTED PRACTICE: XX. SOURCES OF INTERFERENCE ASSOCIATED WITH DIFFERENCES IN LEARNING AND RETENTION[1]

BENTON J. UNDERWOOD AND RUDOLPH W. SCHULZ

Northwestern University *State University of Iowa*

The present studies deal with learning and retention of paired-associate lists as a function of massed and distributed acquisition. As interlist interference builds up across lists, distributed practice is more and more likely to facilitate paired-associate learning; this fact is quite clear if learning is carried to a fairly high level (Underwood & Richardson, 1958). However, there is no indication that for varying levels of interlist interference distributed practice will be followed by appreciably better retention than will massed practice (Underwood & Richardson, 1957, 1958). This is contrary to results found with serial learning where distributed practice gives better retention than does massed practice when interlist interference is high (Underwood & Richardson, 1955). It is this contradiction that is the focal point of the present experiments. The matter could, of course, be dropped by assuming that the contradictory results represent consequences of intrinsic and fundamental differences in paired-associate and serial learning. However, in the interest of empirical continuity across tasks, it does not seem appropriate to make such an assumption until more exhaustive analyses have been made.

In the paired-associate studies referred to above, the lists were constructed so that interlist interference was generalized; formal similarity (repeated letters) was not specific either to stimulus or response terms. Furthermore, in order to make these lists comparable to the serial lists which *had* given better retention following distributed than following massed practice, intralist letter duplication was fairly high. The objective of the present experiments was to isolate these various sources of interference. First, intralist similarity was made as low as possible for all lists. Second, in one set of conditions interlist interference was built up among responses. Third, in another set of conditions interlist interference was built up among stimuli. Finally, by using different classes of materials for stimulus terms and response terms, interference between stimuli and responses was largely eliminated in all lists. The empirical question is: as interference builds up across lists, what is the influence of distributed practice on learning and retention?

METHOD

Lists.—Two experiments are involved. In one the lists had adjectives as stimulus terms and nonsense syllables as response terms. This experiment will be referred to as Exp. A-S (adjectives-syllables), or will be said to use A-S lists. In the other experiment the nonsense syllables were used as stimuli and the adjectives as responses. This will be referred to as Exp. S-A. For both experiments the lists were identical as far as particular items and particular pairings were concerned; the A-S lists were simply "turned over" to form the S-A lists.

[1] This work was done under Contract Nonr-1228, Project 154-057 between Northwestern University and the Office of Naval Research.

TABLE 1

LISTS USED IN EXPERIMENT A-S

List 1	List 2	List 3	List 4
coiled—QIV	noted—QAW	foggy—QOZ	missing—QIX
required—JOP	partial—JOH	bragging—JEP	famished—JUW
perfect—LOH	human—LUZ	diverse—LAH	showy—LOZ
aware—BIK	useless—BEM	skillful—BEK	vocal—BUP
massive—FAW	legal—FOV	cautious—FAX	heathen—FEK
open—GIM	shielded—GUP	knightly—GOM	awkward—GOH
sudden—CAZ	precise—CIX	quiet—CEW	double—CEM
wishful—DUX	bridal—DAK	ailing—DAV	ghostly—DUV

Note.—The lists used in Exp. S-A were the same except that the adjectives were the stimulus terms, syllables the response terms.

The four lists used, presented in their A-S version, are shown in Table 1. Each *S* learned all four lists. The intent was to build up the interference across lists among the nonsense syllables. Within a list, no consonant was used more than once. However, the consonants used in the first position for any given list also occur in the first position for all other lists. The same rule holds for consonants in the last position. In some instances syllables across lists have the same first consonant and same last consonant, thus differing only in the vowel; in some instances this is not true. This inconsistency was intentional to avoid the possibility that *S* would induce a general principle concerning the construction of the syllables. All syllables were within the range of 46% to 67% association value as defined by Glaze (1928). The two-syllable adjectives were chosen to have as low interitem similarity as careful inspection procedures allow.

Procedure.—The procedure for the two experiments was identical; they differed only in the nature of the lists. Experiment S-A was designed to build up interlist interference among stimuli; Exp. A-S to build up interlist interference among responses. Each experiment had two subgroups of 48 *S*s each. One of these subgroups learned all four lists under massed conditions, which consisted of a 4-sec. intertrial interval. The other subgroup learned the first and fourth list under distributed conditions, the distribution being a 60-sec. rest interval inserted after each learning trial. Symbol cancellation was used to fill the 60-sec. intervals.

All *S*s were naive as far as formal verbal learning was concerned. On Day 1 each *S* learned a single list, this list being presented for 12 trials after he had been instructed to give as many correct anticipations as possible on each trial. Twenty-four hours after learning the first list (Day 2), *S* was given 5 relearning trials. Then, second and third lists were presented for 12 acquisition trials each. These lists may be considered filler lists, designed to build up interlist interference. On Day 3, *S* was given a fourth list for 12 acquisition trials. On Day 4—24 hr. after learning the fourth list—*S* was given 5 relearning trials on the fourth list. Relearning was always under massed practice.

To summarize: each *S* learned four lists; the first and fourth lists were recalled after 24 hr. One group of *S*s learned the first and fourth list by massed practice, another learned them by distributed practice. The second and third lists were given under massed practice for all *S*s.

Order of the lists was counterbalanced so that each list occurred equally often at each position in the series of four lists. The anticipation method of learning was used with a 2:2-sec. rate of presentation on a Hull-type memory drum. A randomized order of conditions was made up subject to the restriction that each condition occurred with each type of list 48 times. The *S*s were then assigned to the schedule in order of their appearance at the laboratory. Since a constant number of trials was used, no *S* was lost for failure to learn. When an *S* was lost for other reasons, e.g., mechanical failure, failure to return for recall, etc., he was replaced by the next *S* appearing at the laboratory.

RESULTS

Learning.—The basic response measure used was mean total correct responses over the 12 acquisition trials. This measure is used in plotting the results for the first and fourth

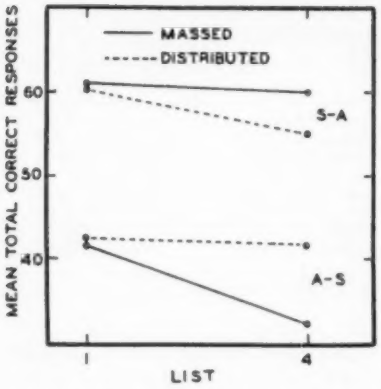

FIG. 1. Learning as a function of massed and distributed practice for syllable-adjective (S-A) lists and adjective-syllable (A-S) lists.

list for both sets of lists in Fig. 1. It may be noted first that the S-A lists were much easier to learn than the A-S lists, which confirms previous findings that meaningfulness of the response term is more critical for learning than is meaningfulness of the stimulus term (e.g., Underwood & Schulz, 1960). Looking at the results for the S-A lists it can be seen that distributed practice produced no effect on learning the first list while for the fourth list it actually resulted in poorer performance than did massed practice. However, statistically speaking, the interaction between lists and intertrial interval was not significant. Thus, as interference increases among stimuli across lists, no facilitation by distributed practice is evident. That there was interference among lists is indicated by the fact that there is no sign of learning-to-learn from the first to the fourth list (thus suggesting increasing interference); this was true in spite of the fact that *S*s were naive to verbal learning at the start of the experiment.

Turning next to the A-S lists, it can be seen that distribution had little influence on learning the first list but it appeared to facilitate the learning of the fourth list. This latter effect is significant statistically. The interaction between intertrial interval and lists gives an F of 6.78, which, with 1 and 94 df, nearly attains the value needed for significance at the 1% level (6.90). The t between mean total correct responses under massing and under distribution on the fourth list is 2.39. It may be concluded that as interlist interference builds up among responses across lists there is an increase in the likelihood that distributed practice will facilitate acquisition. Again, the lack of learning-to-learn is noteworthy.

Retention.—The raw recall scores indicate a heavy increase in proactive inhibition between the first and fourth list. This may be seen by calculating the percentage recall using the number correct on the last original learning trial as the base. For the massed A-S lists the percentage recall for the first list is 78%, for the fourth, 36%. For the S-A lists, the corresponding values are 91% and 44%. However, the critical data concern retention as a function of massing and distribution and it is to this difficult measurement problem that attention is now turned.

As noted earlier, intertrial interval did not influence learning of the first list for either the S-A or A-S lists. Likewise, no significant difference in retention was observed following massed and distributed practice for the first lists. However, evaluation of differences in retention of the fourth list is complicated by the differences in learning as a function of intertrial interval. The learning and relearning curves for the fourth lists are shown in Fig. 2. For the

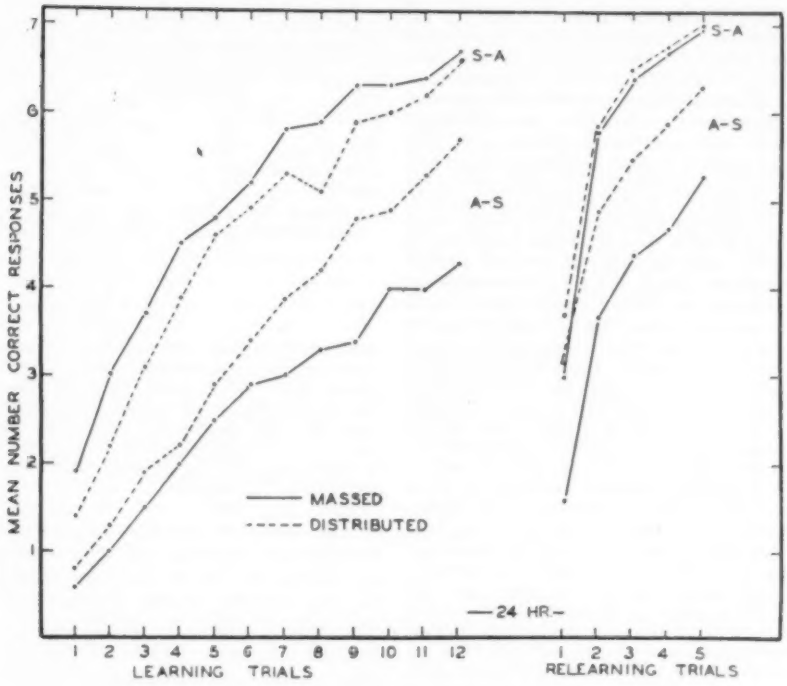

FIG. 2. Acquisition and relearning curves for the fourth lists as a function of massed and distributed learning.

A-S lists, the difference in learning as a function of intertrial interval is clearly apparent. Roughly speaking, however, recall and relearning differences appear to be of the same magnitude as those present at the end of learning. The question is whether or not any difference can be attributed to the method of acquisition (massing vs. distribution) over and above the effect of the methods on level of learning attained. For the S-A lists an evaluative problem is also present. These lists were learned a little better under massing than under distribution, but recall was better following distribution, and the relearning curves show no superiority for the massed group, even though this group was superior throughout learning.

To evaluate differences in recall without confounding by differences in degree of learning, a simple probability analysis has been used to estimate the number of items which would have been correct had a recall trial been given immediately after learning. This method differs somewhat from the one previously described (Underwood, 1954). The first 11 trials were used to determine the reinforcement history of each item. Items were categorized according to the number of times they had been correctly anticipated (reinforced) during the first 11 trials. Thus, each item was classified as having from 0 to 11 reinforcements. Since there were 48 *Ss* and 8 items for each list, a total of 384 items was divided among the 12 categories. For

each category the critical statistic is the proportion of the items which was correct on Trial 12. These proportions were then plotted against number of reinforcements, and extrapolated to determine the proportion which would be correct for items which had been given 12 reinforcements. For the present data little or no error can be present in this extrapolation since the curves are asymptotic at .99 after about 9 reinforcements.

The above procedure yields a table which may be used to indicate the proportion of items which would be correct had another trial been given. For example, consider an item given four times correctly during the 12 acquisition trials. The probability of such an item being correct on a hypothetical next trial is determined by the proportion of items correct on Trial 12 which had been given four times correctly prior to Trial 12. The table of probabilities is, of course, based on the scores of all *S*s combined. However, an expected score can be obtained for each *S*. This method has been explained in detail elsewhere (Richardson & Underwood, 1957). Briefly, each item for a given *S* is assigned a proportion determined by the number of reinforcements that item had during learning. These proportions for the eight items are added together to obtain the expected score for the *S*. A loss score, which is the number of correct responses obtained after 24 hr. subtracted from the number expected, is then obtained for each *S*. These distributions of loss scores can be used for statistical evaluation of differences in retention between lists.

Some evidence should be given about the reliability of the above technique for estimating expected recall. The two scores, predicted for next trial and obtained after 24 hr., can be correlated. For the A-S lists the r for the massed condition on the fourth list is .61, for the distributed, .68. For the S-A lists the corresponding values are .50 and .76. These correlations are highly significant statistically and, in view of the restricted range of recall scores (which would tend to attenuate correlations), they may be taken as indicative of a high relationship between the expected and obtained scores. However, such correlations tell nothing about absolute accuracy of prediction; that is, they do not tell how the mean predicted score for a group of *S*s for a hypothetical next trial would compare with the mean if a trial were actually given. To gather evidence on this matter, predictions have been made for three points in learning the S-A lists for which actual performance was available against which to check the predictions. The results of this analysis show that the median number of items predicted and the obtained median are almost identical. However, the mean number of predicted items tended to be higher than the actual performance means, and this was true for all three tests made. The amount of the discrepancy changes very little as the level of learning increases.

TABLE 2

MEAN LOSS SCORES ON FOURTH LIST FOLLOWING MASSED AND DISTRIBUTED ACQUISITION

Learning Condition	A-S Lists		S-A Lists	
	M	σ_M	M	σ_M
Massed	3.05	.20	3.92	.24
Distributed	2.92	.20	3.05	.21
t	.46		2.72	

The mean loss scores for the fourth lists under massing and distribution are shown in Table 2. The results for the S-A lists indicate that retention was significantly better for the list learned under distributed practice than for the list learned under massed practice. The t (2.72) is significant beyond the .01 level. For the A-S lists distribution during learning does not produce a significant effect on retention.

DISCUSSION

When naive *S*s learned the first list in the present experiments, intertrial interval had no measurable effect on either learning or retention. The intralist similarity was very low. However, as interlist similarity built up across lists, differences in performance as a function of intertrial interval were evident on the fourth list but these performance differences were dependent upon the locus of the interference. The purpose of the present discussion will be to provide tentative explanations of how different loci of interference could mediate the performance differences.

The A-S lists consisted of adjectives

as stimulus terms and nonsense syllables as response terms. As successive lists were learned the interference built up across the nonsense syllables due to repeated letters from list to list. Under these conditions of interference, distributed practice facilitated learning but had no effect on retention. Indeed, the results are quite comparable to those of two previous studies using paired-consonant syllables (Underwood & Richardson, 1957, 1958). In those paired-consonant lists, both intralist and interlist similarity was relatively high but this similarity obtained not only within stimuli and within responses, but also among stimuli and responses. For the S-A lists in the present work, where interference built up among stimuli, distributed practice did not facilitate acquisition. It is possible, therefore, that distributed practice will facilitate learning only when interference obtains in response processes. The present discussion makes such an assumption; but, to specify the locus of this interference with greater particularity is difficult. The reason for this difficulty will become apparent as the discussion proceeds.

In the A-S lists the similarity obtained across responses. If these lists are idealized in terms of a transfer paradigm, they are seen to be of the A-B, C-B′ variety. That is, stimulus terms are quite different, but responses are similar. In a pure A-B, C-B paradigm there is some interference (Twedt & Underwood, 1959) which has been identified tentatively to be a result of backward associations. It is not unreasonable to expect that such interference could become appreciable in amount across four lists even if the responses are only similar (rather than identical). But there is another source of interference in the A-S lists which is much more apparent. As the similarity builds up across lists *S* should experience more and more difficulty in simply learning the responses. The increased similarity across lists should make it more and more difficult to integrate the response term into a readily available unit. It is a fact that complete three-letter interlist intrusions, partial responses, and three-letter responses that did not occur in any list—all of which may indicate difficulty in response integration—occurred with greater frequency in the fourth list than in the first. For the first and fourth lists learned by massed practice, the mean total number of such errors per *S* was 9.6 on the first list and 13.1 on the fourth, although the rate of learning the two lists was essentially the same. Thus, it is possible to identify at least two different sources of interference in the response processes, namely, backward-association interference and integrative interference. Either or both of these may be involved in producing differences in learning by massed vs. distributed practice.

In view of the above considerations it is not possible to specify a single or specific locus of interference responsible for differences in learning by massed and distributed practice. However, even if such specification were possible, the question of how the interference mediates differences in learning by massed and distributed practice must still be considered. Without implying a heavy commitment to the notion, it may be pointed out that an interference-produced inhibition which *retards* the growth of associative strength, and which dissipates very rapidly with rest, will fit most of the facts thus far available. One implication of such an hypothesis for the A-S lists is that there should be a greater number of intralist errors in learning under massed practice than in learning under distributed practice. Thus, if the interference produces an inhibition which retards the growth of associative connection between the stimulus and response term, one consequence of this slower growth should be a greater number of misplaced responses within the list. The data confirm this expectation. In learning the fourth list under massed practice a total of 342 such errors was recorded, while for distributed learning the number was 241. The frequency was greater under massed learning for every trial except the first.

However, other types of errors showed very little difference in frequency, totaling 628 under massing and 659 under distribution.

This evidence on intralist errors can be viewed in a more specific manner. If interference produces an inhibition which retards the growth of associative strength, it would be expected that when *S* first successfully gives a response from the list it would be less likely to be paired with its appropriate stimulus under massed learning than it would under distributed learning. Again, the data support this expectation. Each *S*'s data sheet was searched to determine the first time each response was given (whether correctly paired with its stimulus or not) and then it was noted whether or not the response *was* correctly paired with its stimulus. Under distributed learning, 328 first attempts were recorded of which 36 (11%) were not paired with the correct stimulus. Under massed learning, 284 first attempts were recorded of which 58 (20%) were not paired correctly.

Thus, the data for the A-S lists are consistent with the notion that interference among response processes results in an inhibition which retards the growth of associative connection between the stimulus term and response term. However, as noted earlier, further experiments will be required to specify the exact locus of the interference involved.

Turning to the S-A lists, the explanatory problem changes. For these lists, distributed practice did not facilitate learning but did facilitate retention when interference was high. In these lists, in which syllables were stimuli and adjectives responses, the interference built up among stimuli as the number of lists learned increased. If the similarity relationships are again idealized according to transfer paradigms, these lists will be seen to constitute an A-B, A′-C situation. Stimuli are similar with response terms quite different. Thus, while the stimuli in the fourth list might tend to elicit responses from lists learned previously, there should be no interference in integrating the responses per se, nor should backward associations be a source of interference. While these differences between the S-A and A-S lists may account for the fact that the acquisition varies as a function of distributed practice, they are not obviously associated with the fact that the fourth S-A list was retained better when learned by distributed practice than when learned by massed practice.

The retention data for the S-A lists are consistent with the notion that distributed practice allows for a more permanent extinction of S-R interference tendencies than does massed practice. Recent data (Barnes & Underwood, 1959) are quite convincing regarding the fact that when *S* learns two lists having an A-B, A-C relationship, the A-B association is extinguished or inhibited in some manner. If a recovery process is assumed to occur over time, under distributed practice the extinguished associations will recover more between each trial than they will under massed practice. Such recovered associations would have to be extinguished again. It may be assumed that this alternating recovery and extinction process will lead to a more permanent extinction of the interference tendencies thus facilitating recall following distributed practice. Furthermore, because of greater recovery of error tendencies during distributed practice than during massed practice, learning may actually be inferior under distributed practice. The present results are consistent with this expectation; although the differences are not significant statistically, it is seen in Fig. 2 that performance under distributed practice is inferior to that under massed practice on every trial. It is quite possible that if interference was increased (e.g., using more than four lists) statistically significant differences would result in favor of massed practice.

The successive-extinction hypothesis would lead to certain expectations concerning interlist intrusions during learning and recall. Specifically, more interlist intrusions should occur under distributed learning of the fourth list than

under massed learning, but fewer during relearning following distributed learning than following massed. For learning the error data are consistent with the hypothesis; 11 interlist intrusions occurred during massed learning, 29 during distributed learning. In relearning, 35 errors occurred for each condition. Thus, the intrusions during relearning, while not contradicting the extinction hypothesis, certainly do not give it support.

Summary

The present experiments were designed to narrow down the number of possible sources of interference in paired-associate learning which may be associated with differences in learning by massed and distributed practice, as well as differences in retention following massed and distributed practice.

The four lists learned by each *S* were made up of nonsense syllables and adjectives. For Exp. S-A the stimulus terms were syllables, the response terms adjectives. Interlist and intralist similarity was low among the adjectives. For the syllables, intralist similarity was low but interlist similarity increased as the number of lists learned increased. Forty-eight *S*s learned all four lists by massed practice (4-sec. intertrial interval) while another group of 48 learned the first and last list by distributed practice (60-sec. intertrial interval). Retention of the first and last list was measured after 24 hr. For Exp. A-S the procedure was exactly the same except that the lists had the adjectives as stimuli and the syllables as responses.

The results show that for the S-A lists distributed practice did not facilitate learning for either the first or fourth list, but did enhance retention of the fourth list. This finding was judged to be favorable to a successive-extinction hypothesis.

For the A-S lists, distributed practice facilitated learning of the fourth list, but not the first, and had no effect on retention of either list. This evidence indicated that distributed practice will facilitate acquisition when some minimal amount of interference is present in some aspect of the response processes. It was suggested that the interference produces an inhibition which retards the growth of associative strength.

References

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Richardson, J., & Underwood, B. J. Comparing retention of verbal lists after different rates of acquisition. *J. gen. Psychol.*, 1957, **56**, 187–192.

Twedt, H. M., & Underwood, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

Underwood, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

Underwood, B. J., & Richardson, J. Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists. *J. exp. Psychol.*, 1955, **50**, 39–46.

Underwood, B. J., & Richardson, J. Studies of distributed practice: XVII. Interlist interference and the retention of paired consonant syllables. *J. exp. Psychol.*, 1957, **54**, 274–279.

Underwood, B. J., & Richardson, J. Supplementary report: Interlist interference and the retention of paired consonant syllables. *J. exp. Psychol.*, 1958, **55**, 95–96.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received February 1, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 236–241

STUDIES IN INCIDENTAL LEARNING: IX. A COMPARISON OF THE METHODS OF SUCCESSIVE AND SINGLE RECALLS

LEO POSTMAN AND LAURA W. PHILLIPS

University of California

This paper presents an experimental comparison between the methods of successive recalls and single recalls. In the method of successive recalls the same *S*s are tested at different intervals after the end of learning. In the method of single recalls separate groups of *S*s are used at each retention interval. In the present experiment, the difference between the two procedures was evaluated as a function of the difficulty of the materials and *S*s' set during learning.

It is now clearly recognized that the method of successive recalls yields systematically biased measures of retention (McGeoch & Irion, 1952, 372ff.). A test of recall affords an opportunity for rehearsal. Hence, the performance on successive tests is determined by preceding tests as well as by the conditions of forgetting during the intervening intervals. The available experimental evidence shows that successive recalls may maintain retention at a high level. In the studies of Brown (1923) and Raffel (1934) repeated tests were given for lists of familiar words or names. A high proportion of the items given on an early test were retained on a later test, and some new items appeared in addition. With continuous prose materials, Clark (1940) obtained consistently higher retention scores by the method of successive recalls than by the method of single recalls. As Ammons and Irion (1954) have demonstrated, rises in successive recalls have been erroneously interpreted as evidence for reminiscence because of failure to take account of the bias inherent in the method. Similar methodological problems have arisen in the interpretation of experiments on memory for visual form (Hanawalt, 1937).

While it is clear that successive recalls are not independent of each other, the question arises of whether the effect of a first recall on a second recall is necessarily beneficial. The verbal materials used in past investigations of this problem were either series of familiar words or textual passages. On an immediate test for such materials the proportion of overt errors is not likely to be high. Hence, the responses rehearsed during the first test of recall will be predominantly correct, and such practice will facilitate performance on subsequent tests. A different result may be expected with difficult materials subject to substantial intraserial interference. The proportion of overt errors on the immediate test of recall should increase with the difficulty of the materials. Such errors will be rehearsed and, in the absence of correction, strengthened during the first test. Given a sufficiently high degree of intraserial interference, the net practice effect may be detrimental, rather than beneficial, to performance on a subsequent test of recall. This view receives some support from Rohrer's (1949) finding that rehearsal of a partially learned list of nonsense syllables during an interpolated rest

period leads to a decrement in subsequent recall.

Two considerations suggested that the differential effects of recall on retention might be greater under intentional than incidental conditions. First, intentional *Ss* have been shown to be more subject to intraserial interference than incidental *Ss* in learning nonsense materials (Postman & Adams, 1957). Hence the rehearsal of errors during an initial test should be more detrimental to intentional than incidental learners. Secondly, associations of intermediate strength should benefit more from rehearsal than associations which are either very strong or very weak at the end of learning. There is evidence that responses learned incidentally tend to be either relatively strong or weak whereas associative strengths are more evenly distributed under intentional conditions (Postman, Adams, & Phillips, 1955). When the responses given on the initial test are predominantly correct, as in the recall of adjectives, the beneficial effects of rehearsal should then be greater under intentional than incidental conditions.

Method

General.—The original learning involved either nonsense syllables or adjectives, and either intentional or incidental learning instructions. Under each of these four conditions there was a successive-recall group and a single-recall group. The successive-recall groups were given both an immediate and a delayed test of free recall; the single-recall groups were given only a delayed test. Thus, the total design comprised eight different experimental groups.

Materials.—The learning materials were (*a*) a list of 48 nonsense syllables, and (*b*) a list of 48 adjectives. The nonsense syllables had Glaze (1928) association values between 26.67% and 40.00%. There was, of course, considerable repetition of letters in the list of nonsense syllables. The mean number of repetitions per letter was 5.5 (Range = 2–10). The adjectives were from Haagen's (1948) list and fell within the highest interval on his 5-point scale of familiarity. In this list, the mean frequency of repetitions of first letters was 2.8. There was also some repetition of adjectival endings, such as *ful* and *ant*. There were no synonyms as defined by inclusion in Haagen's word clusters.

Conditions of learning.—The task for all *Ss* consisted of matching each verbal item to a geometric design. The *Ss* were instructed to select one of the following geometric designs as best "fitting" a given item: square, triangle, circle, cross, arrow, diamond, and sine wave. Large-scale drawings of these designs were in full view throughout the learning session. The *Ss* were requested to make their judgments quickly and without attempting to make fanciful connections between verbal items and geometric designs. Each item was exposed for 4 sec. on a projection screen. The height of the letters on the screen was 1.25 in. After exposure of each item, *Ss* were given 6 sec. in which to draw the design they had chosen in the appropriately numbered space of a record blank. New items were, therefore, exposed at a 10-sec. rate. The *Ss* used a cardboard strip to cover all spaces on the record blank except the one in which they were writing. The series was presented twice, each time in a different random order. The interval between successive presentations of the series was 30 sec. The second presentation was stated to be for purposes of studying the variability of matches. The intentional *Ss* were instructed to remember as many of the verbal items as possible. No learning instructions were given to the incidental *Ss*.

Tests of recall.—Five-minute tests of free recall were used to measure retention. The *Ss* were instructed to write down all the items they could remember, in any order that occurred to them, and to include items about which they were not certain. The successive-recall groups were given a first test 30 sec., and a second test 25 min., after the end of learning. During the interval separating the two tests *Ss* worked on a series of mathematical problems. The single-recall groups were tested 25 min. after the end of learning. The *Ss* in these groups spent the entire 25-min. interval working on mathematical problems.

Subjects.—There were 20 *Ss* in each of the eight experimental groups. The *Ss* were assigned to conditions in rotation in the order of their appearance in the laboratory. The procedure was administered to small

groups of *Ss* averaging four in number. On the basis of a postexperimental inquiry incidental *Ss* who had attempted to memorize the lists were discarded and replaced. A total of four *Ss* had to be replaced: one each in the successive-recall groups learning nonsense syllables and adjectives, respectively, and two in the single-recall group learning nonsense syllables.

Results

Immediate recall.—Table 1 shows the mean numbers of correct responses for all recall tests. The first tests of the successive-recall groups provide a measure of retention immediately after the end of learning. Intentional learners recall a significantly larger number of nonsense syllables than do incidental learners ($t = 2.83$, $df = 38$, $P < .01$). For adjectives, the difference is in the same direction but does not reach significance ($t = 1.29$). With highly meaningful materials, marginal differences between the two conditions of learning are to be expected (Postman, Adams, & Phillips, 1955).

Successive recalls.—The change between the first and second tests of the successive-recall groups varies with the nature of the learning materials. The number of nonsense syllables recalled declines for both the intentional and incidental groups. Analysis of variance shows the difference between the first and second successive tests to be significant ($F = 11.11$, $df = 1/38$, $P < .01$). The interaction of Instructions with Tests does not reach significance ($F = 2.38$). The number of adjectives recalled, on the other hand, does not decline significantly between the first and second tests ($F < 1$). The interaction of Tests with Instructions is again not significant ($F = 1.28$).

The majority of items reproduced on the second tests had been recalled on the first tests. However, new items account for 15.7% and 22.0% of the nonsense syllables recalled on the second test by intentional and incidental *Ss*, respectively. For adjectives, the corresponding percentages are 17.7 and 15.5.

Single recalls.—An unbiased measure of retention is yielded by a comparison between the immediate tests of the successive-recall groups and the delayed tests of the single-recall groups. As Table 1 shows, neither intentional nor incidental learners had any retention loss for nonsense syllables. In fact, there is an apparent rise on the delayed test under incidental conditions, but neither the over-all difference between Tests nor the Tests × Instructions interaction is significant ($F < 1$ in both cases). In contrast to the results for nonsense syllables, there was a significant

TABLE 1

Mean Numbers of Items Correctly Reproduced

Test	Nonsense Syllables				Adjectives			
	Intentional		Incidental		Intentional		Incidental	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Successive recalls								
Immediate	5.45	2.42	3.30	2.24	17.10	4.04	15.55	3.32
Delayed	4.45	2.20	2.95	1.86	17.20	4.65	14.85	3.81
Single recalls (delayed)	5.40	2.92	4.00	2.34	14.15	4.36	13.75	4.38

TABLE 2

MEAN NUMBER OF OVERT ERRORS AT RECALL AND MEAN PERCENTAGES THAT ERRORS ARE OF TOTAL RESPONSES

Test	Nonsense Syllables						Adjectives					
	Intentional			Incidental			Intentional			Incidental		
	Mean	*SD*	%	Mean	*SD*	%	Mean	*SD*	%	Mean	*SD*	%
Successive recalls												
Immediate	10.40	5.68	61.3	7.10	4.55	66.8	3.15	3.35	13.8	1.85	1.35	10.3
Delayed	9.75	5.58	63.5	7.95	5.54	70.5	3.35	2.67	15.8	3.05	3.01	15.8
Single recalls (delayed)	9.60	4.82	61.9	8.30	5.06	65.3	4.45	3.77	22.6	3.50	2.16	21.2

amount of forgetting for adjectives as measured by the difference between the immediate tests and the delayed single tests ($F = 6.54$, $df = 1/76$, $.02 < P < .05$). The interaction of Tests with Instructions is not significant ($F < 1$). In view of the unequal amounts learned, the difference in retention loss between nonsense syllables and adjectives cannot be attributed unequivocally to the nature of the materials.

Interaction of materials and methods of testing.—An immediate test of recall is beneficial to the subsequent retention of adjectives, and detrimental to the retention of nonsense syllables. An analysis of variance was performed on the results of all the delayed tests, after conversion of all scores into percentages of the number of items in the list (48) and an arc-sine transformation of these percentages. The interaction of Materials with the Method of Testing was significant ($F = 6.00$, $df = 1/152$, $.02 < P < .05$). This significant interaction supports the conclusion that the effect of recall on forgetting varies with the difficulty of the materials.

Errors at recall.—The mean numbers of errors given on the various tests of recall are presented in Table 2. The table also shows the mean percentages that overt errors were of the total numbers of responses. On the immediate test of recall for nonsense syllables, intentional *Ss* made a larger absolute number of errors than did incidental *Ss*. The difference barely misses significance at the .05 level ($t = 1.98$, whereas 2.02 is required with 38 *df*). There were no significant changes in the number of errors between the immediate test and either of the delayed tests. While the absolute number of errors is consistently higher for intentional *Ss*, the percentages of errors are comparable under the two conditions of learning. Since the absolute number of errors as well as of correct responses is higher for intentional than incidental *Ss*, whereas the proportions of errors are approximately equal, it is possible that the difference in the amount recalled is at least in part a function of the level of responding under the two conditions.

As compared with the results obtained for nonsense syllables, the absolute frequencies of errors during the recall of adjectives are small, as are the percentages that errors are of the total responses. On the immediate test of recall, intentional *Ss* made more errors than incidental *Ss*, but the difference falls short

of significance ($t = 1.81$, $.05 < P < .10$). The number of errors rises on the delayed tests. The increase in errors is significant for the delayed single tests ($F = 5.17, df = 1/76, .02 < P < .05$), but not for the delayed tests of the successive-recall groups ($F = 3.71$, whereas a ratio of 4.10 is required at the .05 level with 1 and 38 *df*). There were no significant interactions of tests with instructions, i.e., the temporal trends in the frequency of errors were parallel for intentional and incidental learners.

The successive-recall groups frequently made the same errors on the immediate and delayed tests. On the second test for nonsense syllables, 26.7% of the errors made by the intentional group, and 25.2% of the errors made by the incidental group, were repetitions of errors made on the earlier test. For adjectives, the corresponding percentages were 28.4 and 31.2.

Discussion

The results of the experiment support the conclusion that the method of successive recalls yields systematically biased measures of retention. The direction of the bias depends, however, on the characteristics of *Ss'* performance on the tests of retention. During the initial recall of nonsense syllables, more than half of the responses were errors, and the effect of the first test on subsequent retention was detrimental. During the recall of adjectives a high proportion of correct responses was given, and the effect on subsequent retention was beneficial.

The expectation that the differential effects of recall on retention would be greater under intentional than incidental conditions was not borne out. While intentional *Ss* made more errors than incidental *Ss* during the initial recall of nonsense syllables, the ratio of incorrect to correct responses was similar under the two conditions of learning and so apparently were the detrimental effects of the rehearsal of errors. The two groups did not differ significantly on the initial test for adjectives, and it is unlikely that intentional learning resulted in more associations of marginal strength than did incidental learning. With the opportunities for rehearsal provided by the initial test, neither group showed a retention loss for adjectives. Thus, the results of the present experiment, like some of the earlier studies in this series (Postman & Adams, 1956; Postman & Phillips, 1954), fail to show significant differences in retention between intentional and incidental learners.

Summary

This paper presents an experimental comparison between the methods of successive and single recalls. The difference between the two methods was investigated as a function of the difficulty of the materials and *Ss'* set during learning. The learning materials were lists of 48 nonsense syllables or 48 adjectives. The conditions of training were either intentional or incidental. Retention was tested by free recall. The successive-recall groups received an immediate and a delayed test, the single-recall groups were given a delayed test only. For nonsense syllables, the interpolation of an immediate test of recall impaired subsequent retention; for adjectives the effect of the interpolated test was beneficial. The direction of the bias introduced by the method of successive recall thus depends on the difficulty of the materials. The two methods of measuring retention produced parallel effects under intentional and incidental conditions.

References

Ammons, H., & Irion, A. L. A note on the Ballard reminiscence phenomenon. *J. exp. Psychol.*, 1954, **48**, 184–186.

Brown, W. To what extent is memory measured by a single recall? *J. exp. Psychol.*, 1923, **6**, 377–382.

Clark, K. B. Some factors influencing the remembering of prose material. *Arch. Psychol.*, 1940, **35**, 73.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Haagen, C. H. Synonymity, vividness, familiarity and association value ratings

of 400 pairs of common adjectives. *J. Psychol.*, 1948, **27,** 453–463.

HANAWALT, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol.*, 1937, **31,** 89.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: III. Interserial interference. *J. exp. Psychol.*, 1956, **51,** 323–328.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: VI. Intraserial interference. *J. exp. Psychol.*, 1957, **54,** 153–167.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. exp. Psychol.*, 1955, **49,** 1–10.

RAFFEL, G. The effect of recall on forgetting. *J. exp. Psychol.*, 1934, **17,** 828–838.

ROHRER, J. H. Factors influencing the occurrence of reminiscence: Attempted formal rehearsal during the interpolated period. *J. exp. Psychol.*, 1949, **39,** 484–491.

(Received February 1, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 242–244

THE EFFECT OF SPEED-UP INSTRUCTIONS ON SPATIAL STIMULUS GENERALIZATION [1]

BERNARD W. HARLESTON

Tufts University

Several studies have demonstrated that spatial generalization can be obtained "with visual stimuli whose position in space is systematically varied" (Brown, Bilodeau, & Baron, 1951, p. 53), and have investigated the role of variables such as the psychological state and chronological age of *S* in determining the shape of the generalization function. Studies in which generalization gradients were obtained have been reported by Brown et al. (1951), Mednick (1955), Bilodeau, Brown, and Meryman (1956), Mednick and Lehtinen (1957), and Arnhoff (1959).

The experiments cited adhered to the same general procedure of training *Ss* to respond to the center lamp of a series of equally spaced lamps and then testing for responses to the other lamps. In the present experiment two procedural changes were introduced to determine their effects on the previously reported generalization functions. One change involved using only one test light for a given *S*; the other change involved training half the *Ss* under repeated pressure to react quickly.

METHOD

Subjects.—Sixty volunteer *Ss* (57 undergraduate students and 3 graduate students) were used.

Apparatus and procedure.—The apparatus was modeled after Mednick's (Mednick & Lehtinen, 1957). It consisted of a horizontal row of 11 lamps, uniformly spaced. Lamp 6 served as the training lamp and Lamps 1, 3, 5, 7, 9, and 11 served as the test lamps. After being told that the task was a test of reaction time and that they should respond as quickly as possible only to the lighting of the center lamp, all *Ss* were given 30 training trials to the center lamp. However, *Ss* receiving the speed-up instructions (the pressured *Ss*) beginning with Trial 11 were urged repeatedly to react quickly and to try to beat their previous score. The *E* applied pressure by saying to *S* after each of 15 of the remaining 20 trials, "Good, now let's see if you can make the next response faster. Good and fast, now. Let's really set a new record." This prompting continued for 60% of the training trials which were interspersed among the test trials. However, *S* was never pressured on the trial preceding the presentation of the test lamp.

The training trials were followed without interruption by a test series in which a given test lamp was presented a total of five times. These presentations were interspersed among 37 additional trials in responding to the center lamp. Five pressured and five nonpressured *Ss* were tested on a given test lamp.

RESULTS

In order to determine whether the speed-up instructions during training affected reaction times, the mean latency scores for the first 10 training trials and the next 20 training trials were compared for the pressured and nonpressured groups. While no differences were found for the first 10 trials, the mean latency score of the pressured *Ss* on the next 20 trials was significantly smaller than that of the nonpressured *Ss* ($t = 4.02$, $df = 58$, $P < .001$).

The number of responses made by the pressured and nonpressured *Ss* to each of the test lamps on each test trial is presented in Table 1.

[1] Portions of this paper were presented at the Washington, D. C. meeting of the American Psychological Association in September 1958.

TABLE 1

NUMBER OF RESPONSES TO EACH TEST LAMP ON EACH TEST TRIAL BY PRESSURED (P) AND NONPRESSURED (N-P) *Ss*

Test Trial	Groups	Test Lamps						
		1	3	5	7	9	11	Total
1	P	2	4	5	5	5	2	23
	N-P	1	4	5	4	3	0	17
2	P	2	4	3	2	2	3	16
	N-P	2	0	4	3	1	3	13
3	P	2	3	3	4	3	2	17
	N-P	2	1	0	1	2	0	6
4	P	3	1	3	4	2	1	14
	N-P	2	0	2	1	0	1	6
5	P	1	0	1	4	3	2	11
	N-P	0	2	3	1	3	2	11
Total	P	10	12	15	19	15	10	81
	N-P	7	7	14	10	9	6	53

When the frequencies are summed over all trials, it becomes apparent that (*a*) the number of responses decreased as the distance between training and test lamps increased, and (*b*) the pressured *Ss* made more responses at each test lamp position. The graphic presentation of these data in Fig. 1 reveals the shape of the generalization functions for the two groups. The differences between pressured and nonpressured *Ss* in total number of responses were significant ($F = 7.96$, $df = 1/48$, $P = .01$).

Trial to trial comparisons across all test lamps reveal that the greatest number of responses was made on the first test trial by both groups. Except for a reversal by the nonpressured *Ss* on the fifth test trial, frequency of responding decreased as the number of test trials increased.

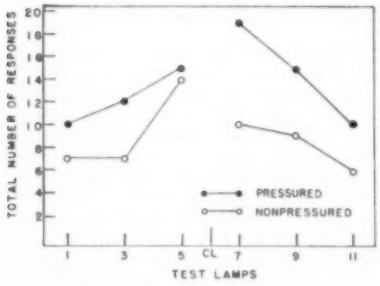

FIG. 1. Relationship between total number of responses to the test lamps on five test trials and the position of these lamps relative to center lamp (CL).

DISCUSSION

The obtained gradients compare favorably with those reported in earlier studies. Further, since a given *S* was tested on only one test lamp, the results indicate that such generalization functions can be demonstrated with independent testing at each test position. In view of Mednick's (Mednick & Freedman, 1960) recently formulated "unit" hypothesis, these results assume added significance, for they demonstrate that generalization gradients may be obtained over a relatively restricted continuum of spatial separation for which all test stimuli are "one unit of separation away."

A parsimonious interpretation of the results with the speed-up instructions suggests that the instructions served as a source of increased drive, strengthening the response tendencies of the *Ss* (Rosenbaum, 1953). The sharper generalization function of the pressured *Ss* underscores the observation of Brown, Clarke, and Stein (1958, p. 359) that "the *Ss* must be constantly urged to hurry if adequate numbers of false reactions are to be obtained."

SUMMARY

After being trained to react to the center lamp of a set of 11 lamps, independent groups of *Ss* were given five test trials on one of six test lamps mounted on either side of the center lamp. The test trials were interspersed among additional training trials to the center lamp. Half of the *Ss* were trained under repeated pressure to react quickly and to try to beat their earlier score. Generalization gradients were found for both the

pressured and the nonpressured *Ss*. Further, the pressured *Ss* produced more responses at each test lamp position. The relationship of the results to earlier studies was noted.

References

Arnhoff, F. N. Stimulus generalization and anxiety. *J. gen. Psychol.*, 1959, **60,** 131–136.

Bilodeau, E. A., Brown, J. S., & Meryman, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51,** 293–298.

Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41,** 52–61.

Brown, J. S., Clarke, F. S., & Stein, L. A new technique for studying spatial generalization with voluntary responses. *J. exp. Psychol.*, 1958, **55,** 359–362.

Mednick, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51,** 536–542.

Mednick, S. A., & Freedman, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57,** 169–200.

Mednick, S. A., & Lehtinen, L. E. Stimulus generalization as a function of age in children. *J. exp. Psychol.*, 1957, **53,** 180–183.

Rosenbaum, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45,** 35–43.

(Received February 6, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 245-256

RESPONSE SPEED, AMPLITUDE, AND RESISTANCE TO EXTINCTION AS JOINT FUNCTIONS OF WORK AND LENGTH OF BEHAVIOR CHAIN[1]

ROBERT FRANK WEISS[2]

Yale University

Work has generally been studied by varying the force necessary to depress the bar in a Skinner box. Length of behavior chain has been studied by varying the length of straight alleys (Besch & Reynolds, 1958; Fredenburg, 1956; Hull, 1934), and also by the use of fixed-ratio schedules in the bar-pressing situation. As regards dependent variables, straight alley experiments involve discrete trials in which performance is measured by the latency of the response, or its reciprocal, response speed. In most Skinner-box studies, however, the bar is freely available to *S*, and performance is measured in terms of the rate of bar pressing (Applezweig, 1951; Ferster & Skinner, 1957; Maatsch, Adelman, & Denny, 1954; Mowrer & Jones, 1943; Stanley & Aamodt 1954; Trotter, 1956). Removing the bar after each trial and re-exposing it at the beginning of the next discrete trial allows one to measure latency, and, *if behavior chains of greater length than a single press are to be studied, the time to complete each press in the chain.* The speed measure was chosen, in part, to permit the data to be interpreted in terms of Hullian theory, which is specified for speed, but not for response rate (Hull, 1943, 1952; Spence, 1956). Moreover, measurement at a relatively molecular level permits the combination of measures which vary similarly into empirically appropriate molar units. Three such units are discussed in this study; starting speed (1st press), pressing speed (following *N* presses), and releasing speed.

A second dependent variable, for which the Hullian theory of instrumental reward conditioning is specified, is the number of unreinforced responses made before *S* meets a criterion of extinction. Though response amplitude is related to excitatory potential (E) only in classical conditioning (Hull, 1952; Spence, 1956), it may be possible to relate instrumental amplitude to inhibitory potential (I) (Weiss, 1958).

The present experiment thus seeks to determine the joint effects of work and length of behavior chain on several kinds of speed measures (starting speed, pressing speed, goal gradient of pressing speeds, releasing speed, total speed), on instrumental response amplitude (including force-on-bar, frequency of incomplete presses, etc.), and on number of responses to a criterion of extinction. Space limitations allow only brief theoretical discussion, later in this report, and also preclude a proper review of

[1] This study is based upon a dissertation submitted to the faculty of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express his gratitude to Frank A. Logan for his cogent advice and generous aid in all stages of the development of this paper. Thanks are also due to Neal E. Miller and Robert P. Abelson for their valuable advice, and to Arlo K. Myers who designed the work-regulating and amplitude-measuring devices. The research was supported, in part, by a grant to Frank A. Logan from the National Science Foundation.

[2] Now at the Research Division, Columbus Psychiatric Institute, Ohio State University.

the literature. The following points may be noted here. There have been no studies of the interaction effects of work and response length on any of the dependent variables of this study. Parametric studies of work effects on any of these variables are presently unavailable, and the range of work levels reported is approximately half that employed here. Work effects have not been studied with the variety of speed measures employed here, nor are work effects on speed in instrumental conditioning (Aiken, 1957) and selective learning (Montgomery, 1951; Thompson, 1944) in clear agreement. Moreover, there have been no studies of response length in which work per unit length is high, and in which response length might therefore affect performance primarily through work duration, rather than through within-chain delay of reinforcement.

Method

Subjects and Design

Five levels of work (15, 50, 100, 150, and 200 gm.) and three levels of response length (1, 4, and 8 presses) were employed in a complete factorial design. The *S*s were 90 male albino rats, 120–150 days old. There were thus 6 rats in each of 15 experimental conditions, combining to make 18 in each work condition and 30 in each length condition. The training took place in 6 separate replications of 15 rats each—one rat to each experimental condition. Four *S*s died and were replaced in later replications (15 gm.—1 press, 100 gm.—1 press, 100 gm.—4 presses, 200 gm.—8 presses).

Apparatus

The apparatus was a modified Skinner box, placed within two concentric soundproofed boxes. The Skinner box proper measured 1 ft. in all internal dimensions. The side of the box on which the bar was located was of aluminum, as were the opposite wall and the hinged ceiling. The remaining two sides and the floor were of ½-in. hardware cloth. The bar was located inside a small well, behind an aluminum door which could be lowered to allow *S* to insert his head and shoulders and press the bar. The bar-well measured 4 in. high by 2 in. wide, and the bottom of the opening was 4 in. from the floor of the box. The bar was T shaped, consisting of an aluminum cylinder 2 in. long and ¼ in. in diameter. The bar was placed ¾ in. in from the mouth of the well, and 4½ in. up from the floor of the box. The magazine delivered approximately .01 cc. of distilled water, by means of a solenoid valve, when the bar was pressed and then released.

The inner soundproofed box was illuminated by two shielded 7-w. bulbs and contained an 8-in. speaker, connected with an oscillator placed outside the boxes. A weak 1000-cycle tone sounded during the 3 sec. before the bar-well door was opened. This box also contained a relay which emitted one dull but audible click when the required force was exerted on the bar, and one somewhat different click when the force exerted dropped below that point. The purpose of this relay was to provide *S* with some feedback in addition to that provided by proprioceptive cues.

The *S* could be observed by mirrors through one-way vision glass. A continuous kymographic recording of amplitude was made, and times were recorded (to 1/100 sec.) by nine Standard Electric timers. The amount of force on the bar (work) necessary to actuate the reward mechanism and the clocks could be varied from 15–200 gm. by varying the distance through which *S* was required to press the bar against a spring resistance.[3] The work-regulating system also included a (constant) waterbath and paddle arrangement which increased the inertia of the bar above that imposed by the spring.

Procedure

Maintenance schedule.—Each *S* received 15 cc. of water in his home cage, once every 24 hr., about 20–30 min. after being removed from the apparatus. Purina laboratory chow checkers were available ad lib. The *S*s were placed on maintenance schedule 15 days before the beginning of training, and were well tamed when training began.

Magazine training.—On Day 1 *S*s were given 25–30 reinforcements at intervals of about 2 sec., whenever they were at the

[3] For example, by 200 gm. is meant that a force equal to a static weight of 200 gm., suspended from the bar, was required to operate the reward mechanism and clocks.

water cup. On Day 2 *Ss* were trained to approach the water cup when the magazine click sounded, each *S* receiving 25–30 reinforcements.

Initial shaping.—On Day 3 *Ss* were rewarded for approaching the bar, until they pressed the bar with a minimum force of 15 gm. and were rewarded automatically. The door to the bar remained open continuously for five rewarded free-responding presses. Following this, *Ss* were given three discrete trials (i.e., the door to the bar was opened at the beginning of the trial and closed when the reward was delivered). All further training used the discrete-trials procedure. The intertrial interval on these three trials was 20 sec. Any *S* that did not make his first bar press within 20 min., or complete initial shaping within 35 min. was discarded. Following initial shaping, *Ss* were divided into three matched groups on the basis of starting speed (1/latency) on the three discrete trials. The three groups were 15 gm., 50 gm., and those being trained to 100 gm. or more. The mean speeds were .076, .077, and .075, respectively.

The following two rules regarding failure to respond with the required force were applied uniformly throughout shaping, acquisition, and extinction. A failure trial was one on which (*a*) latency of the first press was greater than 90 sec., or (*b*) latency of any one of the later presses in a trial was greater than 60 sec. A failure trial was always followed by a make-up trial, after the regular intertrial interval. During shaping and the first day of acquisition, failure on two successive trials (one regular, one make-up) was followed by retraining at the previous shaping stage, and, if necessary, earlier stages.

Work shaping.—With the exception of the 15 gm. group, which had already completed its three discrete trials at its criterion work level, all *Ss* received further work shaping, in blocks of 8 or 16 trials, according to the schedule shown in Table 1. The intertrial interval was 20 sec. After completion of the first day of work shaping, *Ss* which had been trained through 100 gm. were divided into three groups: 100 gm., 150 gm., and 200 gm. Starting speeds for the 16 work-shaping and three initial-shaping trials were combined as a matching variable. Mean speeds of the three groups were .474, .482, and .473.

Response length shaping.—On the day following the completion of his work shaping, each *S* was assigned to a response length condition using starting speed of all available trials. The *Ss* were thus divided into length groups matched within each of the five work groups. The *Ss* assigned to the 4- and 8-press conditions received shaping according to the schedule indicated in Table 1. *All Ss who who completed initial shaping,* and were assigned to groups, *also completed work and length shaping. There was no self-selection of Ss in shaping.*

TABLE 1

SHAPING SCHEDULE

Work Shaping Schedule								
Work in Grams	30	50	75	100	125	150	175	200
Trials at Work Level	5	3	5	3	5	3	5	3

Response-Length Shaping Schedule								
Length in Presses	1	2	3	4	5	6	7	8
Trials at Length Level	1	3	1	3	1	3	1	3

Acquisition.—Acquisition began the day following completion of shaping. All *Ss* received 16 reinforced trials a day for 6 days, a total of 96 trials. The intertrial interval was 40 sec.

A failure trial was always followed by a make-up trial so that there were always 16 reinforced trials in a block. In a few cases *Ss* also failed to complete the make-up trial. When this happened after the first day of acquisition, *S* was removed from the apparatus, and was allowed to complete the remaining trials on the following day. If *S* failed to complete these trials he was discarded. Four *Ss*, 200-gm.—4 press, a 150-gm. —8 press, and two 15-gm.—8 press rats met this criterion of two successive failures on two successive days and were discarded. These *Ss* were replaced in later replications, and the shaping trial speeds of the replacements were used in matching. Some opportunity for self-selection may thus have been present, with respect to the length variable, in 4 out of 90 cases. This opportunity was minimized by the matching procedure.

Extinction and spontaneous recovery.—On the day following the final block of rewarded trials, *Ss* were extinguished in one session to a criterion of two successive failure trials. The *only* differences between acquisition and extinction procedures were that the magazine did not deliver the reward or click at the end of a trial, and that extinction was not divided into blocks of 16 trials. On the following day *Ss* were re-extinguished, using the same procedure. While all *Ss* had the same number

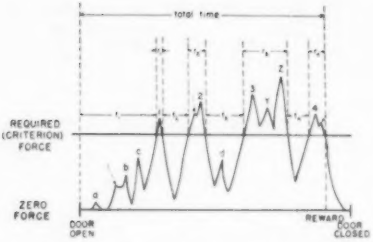

FIG. 1. Hypothetical kymograph record of a single trial for a 4-press chain.

of acquisition trials at their assigned work and response length levels, they had different numbers of shaping trials. Since this difference might affect resistance to extinction, the following control was instituted. In half the replications *S*s were extinguished the day following their 96th acquisition trial. In the other three replications, each *S* received a number of additional reinforced trials equal to the difference between the number of shaping trials he had received, and the largest number of shaping trials given.[4]

Speed scores.—Figure 1 depicts a hypothetical record of the behavior of a rat in the 4-press condition, as it would be recorded by the kymograph. Although time measures were recorded by clocks, rather than being obtained from the kymograph records, the diagram depicts time as well as amplitude measures.

The time from the opening of the bar-well door to the delivery of the reward is called total time. This measure, as shown in Fig. 1, may be divided into the time to make each successive press with the required (criterion) force (t_1, t_2, t_3, t_4) and the time involved in releasing the bar below the required force level after each press (r_1, r_2, r_3, r_4). The time for the first press (t_1) is called starting time, and that for the remaining presses ($t_2+t_3+t_4$) pressing time, this being analogous to the distinction between starting and running time in a straight alley. Releasing time was not recorded for each press separately, but only summed over the 1, 4, or 8 presses in a trial.

The following tabulational procedure was used on all speed data. First speeds were obtained by taking the reciprocal of each latency. Then the mean of these speeds on a given press was computed for each *S* over a block of 16 trials. These individual *S* means then constituted the "raw data" on which group means and statistical analyses were based. On failure trials, the press on which the failure occurred was scored as zero speed. Any presses preceding the failure press on a failure trial were scored as usual. When the failure criterion time for a press was exceeded, the bar-well door was closed. Since *S* was given no opportunity to make the remaining presses on that trial, he was not scored for them. Make-up trials were also unscored.

In straight alley studies, it is customary to correct for differences in alley length by multiplying running speed by the length of the alley. An analogous correction for differences in response chain lengths was made by multiplying speed by the number of presses in the response chain aggregated as a single measure. It was thus possible to determine whether average speed of responding varied with response length.

Amplitude scores.—For purposes of analysis it was necessary to adopt criteria for separating the continuous (within a single trial) kymograph record into separate presses. This was simple in the case of criterion presses. A *criterion press* is a press which stopped a clock (Fig. 1, Presses 1, 2, 3, 4). As can be seen, a criterion press begins below the required amplitude, and reaches its peak above the required amplitude. A *subcriterion press* begins below the required amplitude and reaches its peak below the required amplitude (Fig. 1, a, b, c, d). When the bar was completely released, 5 gm. was the minimal amplitude which counted as a subcriterion press. *Partial release* was defined as a 10% (or greater) decrement in amplitude, and a 10% increment in amplitude, from the level reached after the release, was required for the change to count as a subcriterion press. In Fig. 1 dotted arrows indicate uncounted trivial changes. These criteria also make possible the discrimination of a third type of press. A *supercriterion press* is a press which begins above the required amplitude and reaches its peak above the required amplitude (Fig. 1, Y and Z). Such a press may be said to occur when *S* makes a criterion press, partially releases the bar, and then presses again before dropping below the criterion force level.

The amplitude of a criterion press was measured from a baseline of zero force to the

[4] All overlearning was in blocks of 16. Where the above outlined procedure would result in giving *S* a block with only 8 trials in it, the assigned overlearning trials were rounded to 16s by increasing the number of overlearning trials by 8 on one replication and decreasing by 8 in two.

peak, indicated by the numerals 1–4 in Fig. 1. As previously indicated, only a press which starts below the required force and reaches to or above the required force (criterion press) can stop a clock. Hence the nature of the apparatus makes the division into criterion, subcriterion and supercriterion a natural one.

Results

The results will be presented in three sections: speed, amplitude, and reactions to extinction. Asymptotic data will receive the most complete scrutiny, graphs and statistical tests not being presented in connection with acquisition and extinction curves.

Speed

Test for asymptote.—Since half the *S*s were given additional trials in order to control for the possible effects of overlearning on extinction, it was possible to test for any further change in speed after Trials 81–96. A *t*-test (mean of the differences) indicated a slight, nonsignificant decrease in starting speed from Trials 81–96 to 97–112 ($t = .343$, $P > .80$). There were, moreover, no systematic differences among the various work and response length conditions. An additional *t* test was made, using the total speed measure. The mean of the differences between total speeds on Trials 81–96 and 97–112 was $-.00051$, ($t = .039$, $P > .90$). Speed curves may thus reasonably be considered to be asymptotic on Trials 81–96.

Starting speed.—Figure 2 shows an interaction between work and length of behavior chain: Starting speed was a decreasing function of work, only at 4 and 8 units of chain length, and starting speed decreased as chain length increased only at the higher levels of work. A Work × Length × Replications variance analysis of these data was conducted. Neither replications nor its interaction with other variables were significant when tested against the second-order interaction, the *F*s for these effects all being less than 1. The Work × Length interaction was significant at the .05 level ($F = 2.06$, $df = 8/70$) when tested against the pooled error term (second-order interaction plus the two nonsignificant interactions). The main effect of work was significant ($F = 4.03$, $df = 4/70$, $P < .01$) as was the response length effect ($F = 10.09$, $df = 2/70$, $P < .001$). Aiken's (1957) finding of no asymptotic latency differences, using a 5 vs. 32 gm., 1-unit response chain, panel pushing task, is in accord with these results. During the course of acquisition, the curves of the three chain-length groups steadily diverged from a common starting point, indicating a multiplicative relationship between length of behavior chain and number of reinforced trials.[5] (see Fig. A).

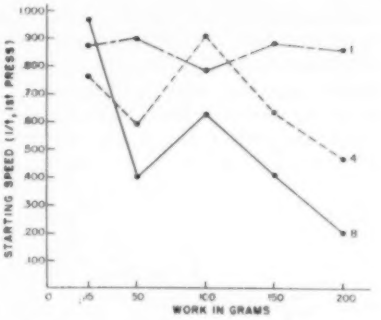

FIG. 2. Asymptotic mean starting speed as a function of work at three levels of length of behavior chain.

[5] Supplementary Fig. A–P have been deposited with the American Documentation Institute. Order Document No. 6563 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Pressing speed.—Figure 3 shows a striking interaction between work and response length. For the 8-press chain pressing speed decreased as work increased, while for the 4-press condition there was a distinct *increase* in pressing speed with increasing work. In order to test not only the Work × Response Length interaction, but also the significance of the work effect at each level of response length, the F test for a linear trend was used. The test for a linear interaction yielded an F of 7.43, which was significant at $<.01$ ($df = 1/54$). The increase in speed as a function of work in the 4-press condition had a linear trend significant at $<.05$ ($F = 4.79$, $df = 1/25$), and the decrease in speed as a function of work at 8 units was also significant at $<.05$ ($F = 4.44$, $df = 1/25$).

Goal gradient of speed.—Goal gradients in this Skinner box situation closely resembled those obtained in straight alleys (Fig. 4). Such gradients typically show an increase, followed by a decrease in speed, the peak generally occurring before the middle of the chain (e.g., Hull, 1934; Weiss, 1960). Repeated-measurements analyses of variance were carried out over Presses 2–4 of the 4-press group and Presses 2–8 of the 8-press group, in order to ascertain whether there was a significant gradient effect after the first sharp increase. This gradient effect was significant for both the 4-press group ($F = 4.86$, $df = 2/50$, $P < .025$) and the 8-press group ($F = 6.73$, $df = 6/150$, $P < .001$) when tested against the within-*Ss* error term. The variance analyses across the second through last presses did not indicate a significant Work × Gradient interaction in either length condition, though in the 4-press chains the peak speed appeared to occur earlier at low work levels than at high ones. The similarity of the 4-unit gradient to the first four trials of the 8-unit gradient is in agreement with Hull's (1934) study of the goal gradient of speed as a function of alley length. "The gradient of a 20 ft. runway has the appearance of a fore-shortened gradient from a 40 ft. runway" (Hull, 1934, p. 419).[6]

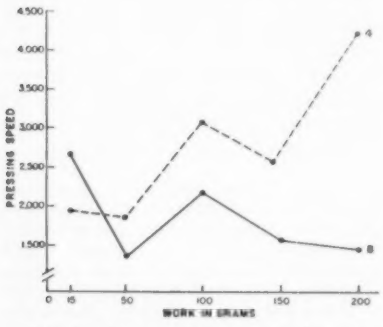

FIG. 3. Asymptotic mean pressing speed as a function of work at two levels of length of behavior chain.

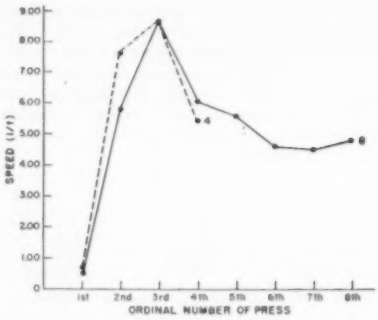

FIG. 4. Goal gradients of mean asymptotic speed at two levels of length of behavior chain.

[6] In Skinner's (1938) early studies of fixed ratio, rate of bar pressing was found to increase as the rat approached the end of the ratio. In Ferster and Skinner's (1957) recent extensive report on reinforcement schedules no changes in local rate are reported at short ratios (apart from the pause immediately following reinforcement), though initial acceleration in the rate of key pecking may occur at ratios over 40:1 if the response has not completely stabilized.

In contrast to Hull's (1934) straight-alley data, the goal gradient became increasingly peaked during the course of acquisition and was flattened by extinction (see Fig. B, C). Straight alley acquisition data recently obtained by the writer (Weiss, 1960) also show the increasing peakedness effect obtained in the present study.

Releasing speed.—At each level of chain length, mean releasing speed increased as work increased. The over-all means, beginning with the 15 gm. level, were 2.97, 2.79, 3.84, 4.86, 6.86. Releasing speed was generally higher for the single-press, than for the 4- or 8-press chains, the over-all means being 5.58, 3.94, and 3.27, respectively. A Work × Length × Replications analysis of variance found both the Work effect ($F = 10.17$, $df = 4/40$) and the Length effect ($F = 8.96$, $df = 2/40$) to be significant at $P < .001$ when tested against the double interaction. In contradistinction to starting and pressing speed, there was no significant interaction effect ($F = 1.60$, $df = 8/40$) (see Fig. E).

Total speed.—It has been shown that work and response length had very different effects on the various molecular components of total time. The values of work and length of behavior chain employed here resulted in starting, pressing and releasing speed effects which completely canceled each other when combined in the more molar total speed measure. A Work × Length × Replications variance analysis of total speed (corrected for chain length) indicated that neither the Work ($F = .73$), Length ($F = 1.40$, $df = 2/40$) nor Work×Length ($F = 1.08$, $df = 8/40$) effects begin to approach statistical significance. It may be noted that total speed was the measure most nearly akin to response rate available in this study, and that these results therefore offer some explanation for the failure of Maatsch, Adelman, and Denny's (1954) study to find significant work effects on asymptotic response rate.

Amplitude

Test for asymptote.—Instrumental response amplitude has not been studied sufficiently to determine whether such measures do, as a rule, reach an asymptote rather than a temporary plateau. By "asymptotic amplitude data" we shall, therefore, mean amplitude measures on Trials 81–96, where speed was asymptotic. Most of the amplitude acquisition curves do, however, appear to have reached some fairly stable level by Trials 81–96. A t test over all 90 *Ss* indicated a non-significant increase in mean amplitude from Trials 65–80 to 81–96 ($t = .48$, $P > .60$).[7]

Frequency of subcriterion presses.—Asymptotic subcriterion press frequency was a multiplicative increasing function of work and length of behavior chain (Fig. 5).[8] Since the data were skewed, a square-root transformation was used before carrying out a Work × Length analysis of variance.[9] The Work × Length in-

[7] It will be noted that the absence of high-work, high-response length *Ss* on Trials 97–112 could decrease the validity of a test for asymptote using these trials, rather than increasing it as in the case of the speed data. The above tests are thus tests for asymptote at Trials 65–80, rather than 81–96.

[8] Applezweig (1951), using a free-responding Skinner box, found that the number of subcriterion presses was not a function of work in later acquisition. Applezweig's work conditions ranged from 10 to 50 gm., 1:1 ratio being employed. Since the 1-press curve in Fig. 5 does not show a marked increase from 15 to 50 gm., the two studies are in agreement within the range of work and chain-length values common to both.

[9] No significant main or interactive effects of replication were found in any of the analyses of the speed data. For the sake of simplicity, the succeeding tests will not

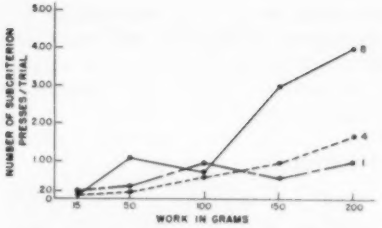

FIG. 5. Asymptotic subcriterion press frequently as a function of work at three levels of length of behavior chain. (Mean number of subcriterion presses per trial preceding the first criterion press.)

teraction was significant ($F = 2.74$, $df = 8/75$, $P < .025$), when tested against the within, as were the effects of work ($F = 17.23$, $df = 4/75$, $P < .001$) and response length ($F = 9.55$, $df = 2/75$, $P < .001$). During the course of acquisition, subcriterion press frequency decreased at 1 and 4 units of response length and at the lower work levels of the 8-press condition. At the highest work levels of the 8-press condition there was an *increase* in subcriterion press frequency during acquisition (see Fig. F and G).

At the higher levels of work there was a marked goal gradient of subcriterion press frequency, the most prominent feature of which was a sharp decrease after the first criterion press, the gradient remaining flat thereafter. The steepness of the gradient was an increasing function of work, the gradient being flat at the lowest level of work. Over-all asymptotic gradient effects yield P values of $<.001$ for both 4-press ($\chi^2 = 68.66$) and 8-press ($\chi^2 = 710.41$) behavior chains. (The null hypothesis assumed equal frequency of subcriterion presses at each point in the gradient.) The two gradients converged from the first to the second press and ran parallel to each other thereafter, the subcriterion press frequency being greater for the 8-press condition at all points. In extinction (mean over-all extinction trials) subcriterion press frequency increased at all points in the 4-press chain and the first 5 presses of the 8-press chain. Both gradients increased in steepness during extinction (see Fig. H, I, J, K, L).

Amplitude of criterion presses.—The median amplitude of the first criterion press on the 16 asymptotic trials was taken for each S, and the required force was subtracted in order to obtain a measure of the amount by which each S exceeded the required force. An examination of these individual S medians indicated that, in the 8-press condition, there was an increasing between-Ss bimodality of distribution with increasing work. A Moses test of extreme reactions, between the 15- and 200-gm. levels of the 8-press condition, gave some statistical support to this observation ($P < .001$). This trend did not appear at the two lower levels of response length.

It appeared that the principal trend of the data could be reasonably well represented by medians, even in the 8-press condition (Fig. 6). Median amplitude was a decreasing function of work at 8 units of chain length, this effect being significant when tested across the entire goal gradient by an above-and-below-the-median chi-square procedure ($\chi^2 = 12.16$, $P < .02$). An analysis of variance on log amplitude for the 4-press condition indicated no effect of Work ($F < 1$). The amplitude of the first criterion press was lower for the single-press chains than for the 4- and 8-press behavior chains (Fig. 6), this difference being significant by an above-and-below medians χ^2 test ($P < .001$).

The asymptotic goal gradient of

evaluate the replications effect separately, their effects being allowed to contribute to the within-cells variance.

criterion press amplitudes was almost perfectly flat, the probability that Gradient effects as great as those observed would arise by chance being $>.99$ for the 4-press chain and $>.80$ for the 8-press chain, as evaluated by an above-and-below the median χ^2 procedure. During the course of acquisition the 4- and 8-press groups increased somewhat in response amplitude, while the 1-press group decreased slightly. The decrement in amplitude during the course of acquisition was most marked and consistent in the extreme low effort (15 gm.—single press) condition. This is in agreement with the results of Marzocco (1951) indicating a decrement in the amplitude of low-work (8-gm.—single press) bar pressing during acquisition, and with similar results recently obtained by Notterman (1959) in a free-responding Skinner box (3 gm.—1:1 reinforcement ratio).

Frequency of supercriterion presses.—The mean number of supercriterion presses immediately following the first criterion press was obtained for Trials 81–96 (see Fig. M). At 1 unit of response length the number of supercriterion presses was essentially zero, while at 4 and 8 units, supercriterion press frequency was a decreasing function of work ($F=7.62$, $df = 4/50$, $P < .001$). (Analysis of variance on 4- and 8-press groups only.) Goal gradients of supercriterion press frequency were obtained, falling steeply and regularly for the 4-press chains ($\chi^2=115.14$, $P<.001$) and more shallowly and less regularly for the 8-press chains ($\chi^2 = 36.17$, $P < .001$) (see Fig. N).

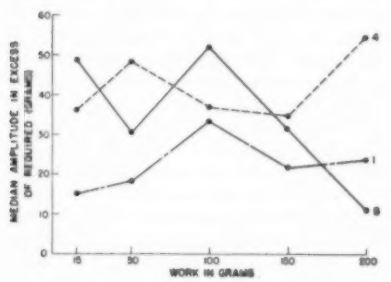

FIG. 6. Asymptotic amplitude as a function of work at three levels of length of behavior chain (median of individual subject median amplitude of the first criterion press in excess of required amplitude).

Trials to Extinction[10]

In testing for the significance of work and response length effects it was necessary to take differences in amount of pretraining into account, by means of the additional "overlearning control" trials given in half the replications. A Work × Length × Overlearning Control analysis of variance was carried out and the main and interactive effects of Overlearning Control were found to be nonsignificant when tested against the variance within cells ($F < 1$ in each case). Any effects of the work and response length variables on extinction are hence not attributable to the pretraining procedures associated with them.[11]

The number of trials to extinction criterion decreased in a negatively accelerated manner as length of behavior chain increased ($F = 18.21$, $df = 2/60$, $P < .001$), the over-all means being 19.0, 10.9, and 10.4, respectively. An inverse relation between work and the number of unreinforced trials to extinction is suggested by the over-all means for

[10] The reader will recall that the term "trial" refers to a complete response chain of 1, 4, or 8 presses at the criterion work level required in acquisition.

[11] Since overlearning control was confounded with replications, the additional precaution was taken of testing for the possibility that there were significant sampling differences between replications. A t test for the difference between the means of overlearning and nonoverlearning replications in asymptotic total speed yielded a t of .084 ($P > .90$).

the five work conditions, which were 16.0, 13.6, 12.2, 14.1, and 11.4. At each level of response length, resistance to extinction appeared greater at 15 and 50 gm. than at 200 gm., but the curves were irregular, and the the variance analysis did not indicate a significant difference among the work groups ($F = 1.88$, $df = 4/60$). An F test for a linear trend, however, approached conventional levels of significance ($F = 3.57$, $df = 1/60$, $P < .07$), leaving the matter somewhat ambiguous. There was no significant Work × Length interaction ($F < 1$).

At each level of response length, the mean number of trials to the second extinction was greater at 15 gm. than at 200 gm., this difference being markedly greater for the 1-press than for the 4- or 8-press chains. While number of trials to second extinction appeared to be a negatively accelerated decreasing function of response length at the 15-gm. level of work, there were no clear differences at higher levels of work. This Work × Length interaction was statistically significant ($F = 4.07$, $df = 8/60$, $P < .001$), but was almost certainly due to "ceiling" effects. The main effects of Work and Length were significant when tested against the within-cells error, the P value for Work being $< .001$ ($F = 6.25$, $df = 4/60$) and that for response Length $< .05$ ($F = 3.84$, $df = 2/60$). (Fig. O, P.)

Discussion

The effects of both work and length of behavior chain appear to be mediated by a work inhibition variable, which may be explained in terms of an underlying conditioned response of stopping-pressing. Asymptotic starting speed was a decreasing function of response length at the higher levels of work only. The absence of such an effect at the lowest level of work appears to indicate that *the effects of response length were due to work-duration, rather than to within-chain delay of reinforcement.*

The work inhibition generated by the work and response length variables may be regarded as having its effect both in terms of simple decrements in performance and also in terms of requiring the animals to choose between patterns of responding involving greater or lesser effort. The simple decremental effect of work is illustrated by the inverse relation between starting speed and work. The increase in releasing speed with increasing work is best understood, however, in terms of choice between patterns of responding involving differential effort. The greater the effort involved in holding down a spring bar, in contrast to that involved in releasing it, the greater the speed with which the bar is released.

The results for pressing speed appear to offer a more complex instance of choice between patterns of responding involving differential effort. Pressing speed was found to be a decreasing function of work in the 8-press condition, and an increasing function of work in the 4-press condition. These results may be interpreted as follows: During acquisition, at response lengths greater than 1-press, the response-produced stimuli resulting from the making of a criterion press become conditioned stimuli for making a succeeding criterion press. A pause between presses has two effects: (*a*) work inhibition decays, facilitating responding; and (*b*) the traces of the response-produced stimuli for making a criterion press also decay, reducing the probability of making a criterion press (generalization decrement) and, therefore, increasing the probability of making one or more subcriterion presses before another criterion press is made.[12] Since the making of subcriterion presses accumulates work inhibition, there are two factors acting in opposite directions: the law of less work, and decrement in temporary work inhibition. Since neither of these factors is adequately quantified, it is impossible to state which is more powerful under any given condition. It is, however, very clearly possible to

[12] In this connection it may be recalled that subcriterion press frequency was found to be greatest following a pause (40 sec. intertrial interval) in pressing.

state the conditions which would increase the potency of the facilitating effects of a pause (or slow pressing) as against its detrimental effects, i.e., any factor, such as massing or response length (work duration), which would increase the accumulated work inhibition. Thus, rapid pressing minimizes work (by minimizing subcriterion presses), but at 8-units of response length the accumulated work inhibition forces pauses or slower pressing to permit its dissipation.

Regular goal gradients of speed were obtained, reaching a peak on the third press in both the 4- and 8-press chains and declining thereafter. The goal gradients thus closely resemble those found in straight alleys, including the terminal decrement in speed, despite the fact that it is not physically necessary for *S* to slow down gradually before stopping in the bar-press situation. These results may be regarded as reflecting, in an unusually pure form, the generalization of a response of stopping-pressing from the end of the chain, where it is learned, to the earlier segments of the chain where it interferes with the bar-pressing response.

The effects of work and response length on the amplitude measures appear, as with the speed measures, to be mediated by a work inhibition variable, based on a stopping-pressing response. The frequency of subcriterion presses may be regarded as an index of the competition between the instrumental pressing response and the inhibitory response of stopping pressing. When the momentary reaction potential of the stopping response exceeds that of the pressing response, anticipatory stopping, i.e., subcriterion presses, should occur. Such anticipatory stopping would be expected to increase with work and response length (work duration), and this expectation was borne out by the subcriterion press results. Median amplitude was a decreasing function of work at the highest level of response length, where increments in work could accumulate sufficient amounts of work inhibition for anticipatory stopping to reduce amplitude of bar-pressing. At the highest combination of work and response length, anticipatory stopping, as reflected in subcriterion press frequency alone, reaches a mean level of 4.04 anticipatory stops preceding the first successful press on each trial. Under such conditions it is not surprising that anticipatory stopping should also cut off presses which have just barely reached the required level, and cut them off so frequently as to significantly decrease median amplitude. It would thus appear that while instrumental response amplitude does not directly reflect excitatory potential (e.g., amplitude does not increase with training), it does bear a determinate relationship to work inhibition.

The effects of work and length of behavior chain on the several speed and amplitude measures were thus in good agreement with a stopping-response theory of work inhibition (Hull, 1943; Miller & Dollard, 1941; Weiss, 1958). The resistance to extinction data were, however, ambiguous in this regard. Of particular interest was the fact that the several relatively molecular measures of speed (starting, pressing, releasing) showed clear results which canceled each other out when combined in a more molar total speed measure.

Summary

Ninety thirsty male albino rats were run in a discrete-trials Skinner box. Within a 5×3 factorial design, there were five levels of force required to be exerted on the bar (15, 50, 100, 150, and 200 gm.) and three levels of length of behavior chain (1, 4, and 8 presses). Sixteen trials (of 1, 4, or 8 presses each) were given each day, for 6 days, the bar being unavailable to *S* for an interval of 40 sec. between trials. Latencies were obtained for each of the *N* presses in a behavior chain, and the time involved in releasing the bar was timed separately from this. Latencies were translated into speeds by the reciprocal transformation. A record of force exerted on the bar (response amplitude) was also obtained. The following conclusions are based on performance at asymptote.

1. Starting speed (speed of first press in chain) was a decreasing function of work at the two higher levels of response length, but was unaffected by work at the 1-press level. Starting speed was a decreasing function of response length at the higher

levels of work only, indicating that the effects of response length were due to work duration, rather than to within-chain delay of reinforcement.

2. Pressing speed (mean speed of presses following the first press in the chain) was a decreasing function of work in the 8-press condition and an increasing function of work in the 4-press condition.

3. Goal gradients of speed were obtained, reaching a peak on the third press in both the 4- and 8-press chains and gradually declining thereafter.

4. Releasing speed was an increasing function of work and a decreasing function of response length.

5. Total speed, the combination of starting, pressing and releasing speed, did not show any effect of work or response length.

6. Amplitude (in excess of force required) was a decreasing function of work at the highest level of response length (work duration).

7. Frequency of subcriterion presses (presses which failed to meet the required work levels and hence did not affect the clocks or reward mechanism) was a multiplicative increasing function of work and response length. Subcriterion press frequency decreased after the first criterion press in a trial, the gradient remaining flat thereafter.

8. The number of unreinforced trials (of 1, 4, or 8 presses) to a criterion of extinction was a decreasing function of response length, and may also possibly ($P < .07$) have been a decreasing function of work. Work and response length effects were both significant in the second extinction.

References

Aiken, E. G. The effort variable in the acquisition, extinction and spontaneous recovery of an instrumental response. *J. exp. Psychol.*, 1957, **53,** 47–51.

Applezweig, M. H. Response potential as a function of effort. *J. comp. physiol. Psychol.*, 1951, **44,** 225–235.

Besch, N. F., & Reynolds, W. F. Alley length and time of food deprivation in instrumental reward learning. *J. exp. Psychol.*, 1958, **56,** 448–452.

Ferster, C. B., & Skinner, B. F. *Schedules of reinforcement.* New York: Appleton-Century-Crofts, 1957.

Fredenburg, N. C. Response strength as a function of alley length and time of deprivation. Unpublished master's thesis, State University of Iowa, 1956.

Hull, C. L. The rat's speed of locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17,** 393–422.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Maatsch, J. L., Adelman, H. M., & Denny, M. R. Effort and resistance to extinction of the bar-pressing response. *J. comp. physiol. Psychol.*, 1954, **47,** 47–50.

Marzocco, F. N. Frustration effect as a function of drive level, habit strength and distribution of trials during extinction. Unpublished doctoral dissertation, State University of Iowa, 1951.

Miller, N. E., & Dollard J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

Montgomery, K. C. An experimental investigation of reactive inhibition and conditioned inhibition. *J. exp. Psychol.*, 1951, **41,** 29–51.

Mowrer, O. H., & Jones, H. M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33,** 269–386.

Notterman, J. M. Force emission during bar pressing. *J. exp. Psychol.*, 1959, **58,** 341–347.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Stanley, W. C., & Aamodt, M. S. Force of responding during extinction as a function of force requirement during conditioning. *J. comp. physiol. Psychol.*, 1954, **47,** 462–464.

Thompson, M. E. Learning as a function of the absolute and relative amounts of work. *J. exp. Psychol.*, 1944, **35,** 506–515.

Trotter, J. R. The physical properties of bar-pressing and the problem of reactive inhibition. *Quart. J. exp. Psychol.*, 1956, **8,** 97–106.

Weiss, R. F. Response speed, amplitude, and resistance to extinction as joint functions of work and reinforcement ratio. Unpublished doctoral dissertation, Yale University, 1958.

Weiss, R. F. Deprivation and reward effects on speed throughout the goal gradient. *J. exp. Psychol.*, 1960, **60,** 384–390.

(Received February 8, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 257–258

SUPPLEMENTARY REPORT: DIFFERENTIAL POSITION HABITS AND ANXIETY IN CHILDREN AS DETERMINANTS OF PERFORMANCE IN LEARNING[1]

ALFRED CASTANEDA

University of Texas

Castaneda and Lipsitt (1959) found a statistically reliable tendency for time pressures (stress) to facilitate performance when the dominant response was compatible with the to-be-learned response, but to interfere when it was incompatible. The present study employs this same task with two groups of children selected on the basis of scores on the CMAS, the children's form of the Manifest Anxiety Scale (Castaneda, McCandless, & Palermo, 1956).

Method.—Group A (anxious) was composed of 18 *Ss* (10 girls, 8 boys) whose CMAS scores fell within the upper 20% of the original standardization population (25–42) and Group NA (nonanxious) was composed of 18 *Ss* (9 girls, 9 boys) whose scores fell within the lower 20% (0–10). All *Ss* were from Grade 5. No *S* had an *L* scale score of greater than 5.

The apparatus was identical to that of the previous study. Eight lamps were arranged horizontally, 2 in. apart, with a push-type button situated 2 in. directly below each lamp. Reading from left to right and letting the first number represent the horizontal position of the lamp, the second the response button, the correct S-R (lamp-button) combinations were 1–1, 2–3, 3–2, 4–4, 5–5, 6–7, 7–6, and 8–8. Thus, for Lamps 1, 4, 5, and 8 the dominant tendency (to select the button directly underneath) was compatible with the correct response (DTC), but for the remaining lamps it was incompatible (DTI). A randomly selected lamp was activated for 4 sec. and *S* was instructed that failure to respond within this interval was an error (all *Ss* responded within this interval during training). Each light was presented 10 times, once in each block of 8 presentations. All other aspects of the procedure were identical to those reported previously.

Results and discussion.—Percentages of correct responses in successive 20-trial blocks for Groups A and NA are shown in Fig. 1. Anxiety facilitates performance in the case of the DTC combinations but impairs performance in the case of the DTI combinations. These results parallel those observed for the stress group in the previous study. An analysis of variance (Lindquist, 1953, Type VI) of these data yielded three reliable effects: Blocks ($F = 28.96$, $df = 3/102$, $P < .001$), S-R Combination ($F = 127.79$, $df = 1/34$, $P < .001$) and Anxiety × S-R Combination ($F = 6.97$, $df = 1/34$, $P < .025$). The difference between the two groups was significant on the DTI combinations ($F = 4.57$, $df = 1/34$, $P < .05$) but not in the case of the DTC combinations ($F = 3.60$, $df = 1/34$, $P > .05$).

Errors occurring to the DTI combinations were differentiated either as dominant tendency errors (DTE) or random errors (RE). The former consisted of responses to the button directly underneath, the latter of responses to buttons other than the one directly underneath the lamp. The analysis applied to these data (see Fig. 2) was the same as that applied to the correct response

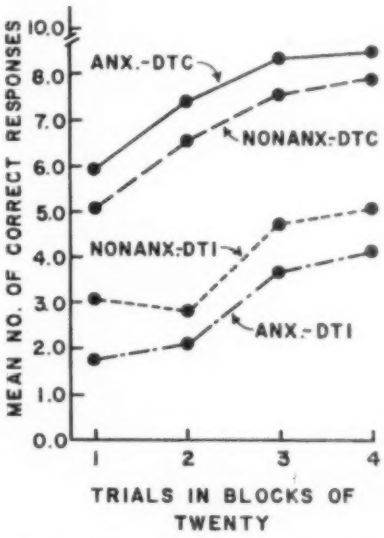

FIG. 1. Mean percentage of correct responses for the two groups on the compatible (DTC) and incompatible (DTI) S-R relations.

[1] The data were collected at the State University of Iowa with the assistance of Lewis P. Lipsitt. The report was supported by a grant from the Hogg Foundation, University of Texas.

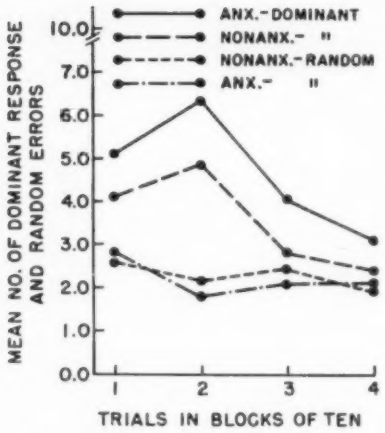

FIG. 2. Mean number of errors attributable to the dominant tendency (DTE) and to random factors (RE) for the two groups.

data. The following effects were found to be reliable: Anxiety ($F = 5.07$, $df = 1/34$, $P < .05$), Blocks ($F = 15.52$, $df = 3/102$, $P < .001$), Type of Error ($F = 41.02$, $df = 1/34$, $P < .001$), Blocks × Type of Error ($F = 66,2$. $df = 3/102$, $P < .001$) and Anxiety × Type of Error ($F = 4.22$, $df = 1/34$, $P < .05$). Group A made more errors of the DTE type ($F = 6.02$, $df = 1/34$, $P < .025$) in comparison to Group NA. The increase in errors of the DTE type from Block 1 to Block 2 was not significant for either group.

In general, these results indicate that stress (time pressures) and anxiety in children (CMAS scores) produce comparable effects in the type of learning problem used in both studies. Both studies are consistent with the theoretical interpretations of Hull and Spence (1956) concerning the role of drive (D) in those forms of learning more complex than that involved in classical and simple instrumental conditioning experiments.

REFERENCES

CASTANEDA, A., & LIPSITT, L. P. Relation of stress and differential position habits to performance in motor learning. *J. exp. Psychol.*, 1959, **57**, 25–30.

CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpm.*, 1956, **27**, 317–326.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received June 9, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 259–260

SUPPLEMENTARY REPORT: EFFECT OF REDUNDANT RELEVANT INFORMATION UPON THE IDENTIFICATION OF CONCEPTS[1]

LYLE E. BOURNE, JR. AND ROBERT C. HAYGOOD

University of Utah

Bourne and Haygood (1959) found a facilitating effect of redundant relevant stimulus information on concept identification performance. The effect was attributed to the fact that increasing the amount of such information, i.e., the number of redundant relevant stimulus dimensions, increases the number of cues *S* can use to identify a set of stimuli correctly. The facilitation occurred at each of three levels of nonredundant irrelevant information, although it became more marked as irrelevant information increased. Rappaport (1957) showed that relevant stimulus redundancy facilitated rapid discrimination of visual forms in the presence of background noise (irrelevant information) but had an inhibiting effect in a noise-free situation. A similar inhibiting effect of relevant redundancy might occur in concept identification if, as Rappaport suggested, *S* attempts to discriminate and use all (redundant) details of each stimulus in making his identification. The present experiment extends the range of irrelevant information used in the previous study (Bourne & Haygood, 1959) and, thus, provides a test of the hypothesis concerning the effects of relevant redundancy in a noise-free task.

Method.—The task, stimulus dimensions, and apparatus used were identical with those of Bourne and Haygood (1959). Seven groups of 15 *S*s from elementary psychology courses performed on a concept identification problem with no irrelevant information and from one to seven relevant dimensions, i.e., levels of relevant redundant information. As before, a problem consisted of a series of geometric stimulus patterns, each of which was a composite of the levels of binary stimulus dimensions. To each pattern, *S* responded by pressing one of two available buttons to identify the category to which the pattern belonged. Since all dimensions were redundant, the levels within all dimensions in any problem were perfectly correlated, e.g., with color and form redundant, squares were always green, triangles red. As relevant information increased, the complexity of each pattern increased although the number of different patterns remained the same. In any problem, *S* was presented with only two discriminably different patterns; he had to learn which of the two responses must be made to each.

Results.—Figure 1 is a composite representation of the data from the present and the previous study. As with one, three, and five irrelevant dimensions, under conditions of no irrelevant information errors to solution decrease as relevant information increases. The results for the groups with no irrelevant dimensions are consistent with the earlier finding that the slope of the performance-relevant information function increases with number of irrelevant dimensions. Number of relevant dimensions had a significant effect on performance in problems with no irrelevant information ($F = 3.58$, $df = 6/98$, $P < .01$). Orthogonal polynomial analysis of this function indicated a linear component only ($F = 15.85$, $df = 1/98$, $P < .01$). This comports with similar analyses on the functions for one and three irrelevant dimensions. An analysis of variance, performed on combined data from all groups with one or two relevant dimensions,

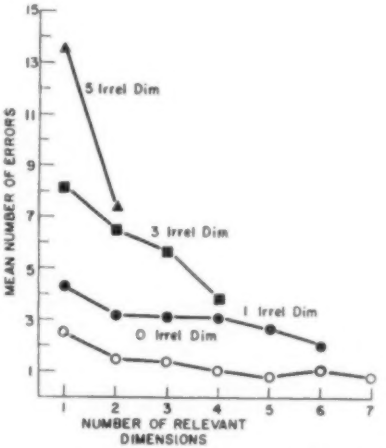

FIG. 1. Mean number of errors to solution as a joint function of number of redundant relevant and nonredundant irrelevant stimulus dimensions. Each plotted point represents the data from 15 *S*s.

[1] This study was supported in part by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service.

TABLE 1

ANALYSIS OF VARIANCE ON ERRORS TO SOLUTION

Source of Variance	*df*	*F*
Number of relevant dimensions (R)	1	20.15*
Number of irrelevant dimensions (I)	3	46.10*
R × I	3	4.85*
Residual	112	(9.43)*

* Number in parentheses is error mean square.
* $P < .01$.

is summarized in Table 1. This analysis included all four levels of irrelevant information, no irrelevant dimensions from the present experiment and one, three, and five from the previous experiment. The significance of the Relevant × Irrelevant Information interaction verified the apparent change in slope of the functions in Fig. 1.

The results of this experiment indicate that relevant redundancy facilitates rather than inhibits performance on a concept identification problem with no irrelevant information. Although it may be argued that such a problem is not noise-free, i.e., that apparatus, instructions, and sets may contribute irrelevant information, the same proposition is no less true for visual discrimination tasks (Rappaport, 1957). It must be concluded that fundamental differences in the nature of the task and the performance required contribute to the differing results.

REFERENCES

BOURNE, L. E., JR., & HAYGOOD, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, **58,** 232–238.

RAPPAPORT, M. The role of redundancy in the discrimination of visual forms. *J. exp. Psychol.*, 1957, **53,** 3–10.

(Received June 13, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 261–262

SUPPLEMENTARY REPORT: THE EFFECTS OF THE MEAN, MIDPOINT, AND MEDIAN UPON ADAPTATION LEVEL IN JUDGMENT[1]

ALLEN PARDUCCI AND LOUISE M. MARSHALL

University of California, Los Angeles

It was previously reported that independent manipulation of either the midpoint (one-half the sum of the two end values) or the median value of a set of stimuli presented for judgment had significant effects upon the centering of the judgment scale, but manipulation of the mean had no systematic effects when midpoint and median were held constant (Parducci, Calfee, Marshall, & Davidson, 1960). These conclusions were based upon experiments in which *Ss* judged the numerical magnitude of numerals printed and exposed together, in order of magnitude, on a single page. The present report extends the previous results to judgments made by the method of single stimuli.

Method.—Each of the six sets of numerals labeled in Table 1 below (and described in Parducci et al., 1960) were judged by different groups of 30 college sophomores. Sets NA and PA differed markedly with respect to midpoints, LL and LH with respect to medians, and PM and QM with respect to means. The preliminary instructions were as follows: "This is a study of how people judge the sizes of different numbers. Before I give you the instructions for judgment, I am going to read you the series of numbers you will later have to judge." After quickly reading aloud the 44 numerals composing the set, *E* gave the previous instructions for judgment (modified to make them appropriate for the new method of presentation) and then read through the set twice again, so that *Ss* heard the entire set three times. The *Ss* were allowed 5 sec. to record their judgments of each numeral. The numerals were read in randomized order, with a different order for each of the six subgroups (of from four to six *Ss*) presented each set of numerals. Thus a varied sample of the previous sets was represented, this time orally and by the method of single stimuli rather than visually with all of the stimuli present while judgments were being made.

Results.—All but one of the new adaptation levels (AL defined as the mean of the numerals judged "medium" on the third exposure to the set), correspond closely to the values predicted by the regression equation derived from the data of the previous experiment (see Table 1). A separate 2×2 analysis of variance was performed upon each pair of new ALs and the corresponding pair of ALs from the previous experiment. The over-all effects of variation in midpoint and median are again statistically significant ($P < .01$ in both cases), while neither the effects of the change in method of presentation nor the interaction between this change and the differences in midpoint or median approach significance ($P > .05$). The low value of the new AL for the PA set appears to be associated with a tendency to omit the "very small" category. Fewer than half the *Ss* judged 351 (the smallest numeral in the PA set) "very small"; the mean AL for these 13 *Ss* was 565, much closer to AL_p and almost identical with the previous AL for the PA set. Since end categories were omitted by only three of the NA *Ss* (and by none of the *Ss* judging the other sets), the basis for this omission remains obscure.

The direction of the difference between the mean ALs for the new PM and QM groups is actually reversed from the direction of the old difference for this pair and also opposite to the direction predicted from adaptation-level theory as interpreted in the previous report. This interaction ($P < .01$) represents the only major deviation from the former results. Comparison of the category-width curves (as in Fig. 2 of the previous report) suggested that this interaction is

TABLE 1

MEAN ADAPTATION LEVELS

Cond.	Old ALs[a]		New ALs		
	AL_p	AL_m	AL	*SD*	% Error[b]
NA	482	494	486	49	.6
PA	581	567	537	51	6.8
LL	501	494	515	53	1.6
LH	546	548	539	73	.8
PM	524	532	549	41	2.8
QM	524	551	520	43	.4

[a] AL_m = AL obtained in previous experiment; AL_p = AL predicted from previous data by the following regression equation: $AL_p = .547(\text{Midpoint}) + .450(\text{Median}) - .027(\text{Range})$.

[b] Difference between AL obtained in present experiment and AL_p, divided by .01 (Range).

[1] This research was supported by Grant G-7485 from the National Science Foundation.

associated with a general decrease, under the new exposure procedure, in the hypothesized tendency to use the respective categories with proportionate frequencies. For each set of numerals, the biggest differences between the old and new curves occur for categories applied to parts of the range in which the differences between successive numerals are either unusually small or unusually large. For parts of the range with closely packed numerals, the categories are wider when exposure is by the method of single stimuli; where the numerals are sparse, the categories are wider using the old method of simultaneous exposure. Thus for the QM set, the "very large" category is wider under the new procedure; for the PM and also for the PA sets, it is wider under the old procedure. Since the distribution of numerals judged "medium" tended to be negatively skewed for the new PM *Ss* and positively skewed for the new QM *Ss*, the difference between their respective ALs would be less if the measure of AL were closer to the midpoint of the "medium" numerals (i.e., closer than their mean which had been selected as a rough approximation to *Ss*' hypothetical compromise between midpoint and median). Similar differences in skewness for the numerals judged "medium" appear to account for the absence of either a significant increase in the difference between the ALs for NA and PA or a significant decrease in the difference between the ALs for LL and LH.

A possible interpretation of the apparent loss of sensitivity to irregularity in the spacing of the stimuli is that single-stimulus *Ss* learned less about the relative frequencies with which they were using the respective judgment categories. This seems reasonable since the random order of presentation insured that the recorded categories would be scrambled rather than conveniently ordered on the response sheet as with the method of simultaneous presentation. But whatever the basis for this loss of sensitivity, the more general features of the results are clear. In spite of the radical change in the method of presenting the stimuli, AL again varied directly with variation in either midpoint or median. And again the effects of variation in mean appeared to be inconsistent with the role attributed to the mean in the theory of adaptation level.

Reference

Parducci, A., Calfee, R. C., Marshall, L. M., & Davidson, L. P. Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *J. exp. Psychol.*, 1960, **60**, 65–77.

(Received September 12, 1960)

Journal of Experimental Psychology
1961, Vol. 61, No. 3, 263–264

SUPPLEMENTARY REPORT: RECOGNITION RESPONSES TO AMBIGUOUS CUES FOLLOWING PAIRED-ASSOCIATE LEARNING

ARNOLD BINDER

Indiana University

Recent experiments (Binder & Feldman, 1960) have demonstrated a close relationship between the relative frequencies of occurrence of figure-name pairs during paired-associate learning trials and the relative frequencies of naming responses to subsequently presented overlapping cues. This confirmed a hypothesis of Binder (1955). In all experiments, the original paired-associate learning proceeded to a criterion. Since certain inferences from the data depended on the assumption of asymptotic learning over a great many trials, it was desirable to determine whether the same relationships would be found if learning involved a fixed number of trials rather than a criterion.

Method.—The stimuli and apparatus were the same as Exp. IV of the series reported by Binder and Feldman (1960), and the general procedure the same as Exp. II. The learning figures are designated C_1, C_2, C_3, and C_4, and the test stimuli C′, C″, and C. C′ contains only cues common to C_1 and C_2, C″ only cues common to C_3 and C_4, and C cues common to all C_i. Each C_i, together with its associated nonsense name, appeared in every learning block. But either C_1 or C_2 occurred four times and either C_3 or C_4 occurred twice, while the remaining one in each pair occurred once (the choice of frequency within pairs was random).

For two groups of 36 *Ss* learning consisted of 10 blocks, followed by a 3-min. rest period and then another 2 blocks. Group C′ *Ss* were then shown test stimulus C′ or C″ (as randomly determined) under instructions to respond with the most appropriate of the previously learned names. Two more learning blocks and a final test trial, involving the remaining one of these two test stimuli, came next for these *Ss*, in that order. A test trial consisting of stimulus C came after the 12 learning blocks for Group C *Ss*. This terminated the experiment for them.

A correction procedure was used on all learning trials. No names were shown in the exposure apparatus during the test trials.

Results and conclusions.—The following notation will be used to categorize test responses according to the learning frequencies per block of the figures designated: C(4) refers to either C_1 or C_2 whichever occurred four times per block for a given *S*; $C(1)_4$ to the one of these two figures which occurred once; C(2) to C_3 or C_4, whichever occurred twice; and $C(1)_2$ to the one of these latter figures which occurred once.

Table 1 shows the obtained frequencies of responses to the various test stimuli. The inappropriate responses are those test responses which are inconsistent with the test stimulus in that they designate learning figures having cues which conflict with the given test stimulus. The relatively few inappropriate responses obtained is consistent with previous results.

The frequencies expected on the basis of the hypothesis that relative frequencies among test responses equal relative frequencies of learning occurrence are shown in parentheses in Table 1. They were computed in the manner described by Binder and Feldman (1960). As previously, moreover, the low-frequency figures were combined to avoid very small expected frequencies.

TABLE 1

OBTAINED AND EXPECTED (IN PARENTHESES) RESPONSE FREQUENCIES TO TEST STIMULI

Test Stimulus C″ (Group C′ *Ss*)		
Appropriate Responses		Inappropriate Responses
C(2)	$C(1)_2$	C_1 or C_2
22	11	3
(22)	(11)	

Test Stimulus C′ (Group C′ *Ss*)		
Appropriate Responses		Inappropriate Responses
C(4)	$C(1)_4$	C_3 or C_4
26	7	3
(26.4)	(6.6)	

Test Stimulus C (Group C *Ss*)		
C(4)	C(2)	C(1) [$C(1)_4$ or $C(1)_2$]
19	6	11
(18)	(9)	(9)

The correspondence between expected and obtained frequencies is as close as could possibly be obtained in the responses to test stimuli C″ and those to C′. While the fit is not as close in the responses to stimulus C, there is little basis for rejecting the hypothesis of equivalence even in this case.

Six *Ss* in the present experiment (three in each group) would not have met the criterion

requirements of Binder and Feldman (1960), but the most general picture was of overlearning, if the previous definition of criterion learning were accepted. Seven or more consecutive perfect blocks were achieved by 58% of the *Ss* over the course of their learning, and 80% achieved four or more perfect learning blocks consecutively.

References

Binder, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62**, 119–129.

Binder, A., & Feldman, S. E. The effects of experimentally controlled experience upon recognition responses. *Psychol. Monogr.*, 1960, **74**(9, Whole No. 496).

(Received October 8, 1960)

PSYCHOLOGICAL MONOGRAPHS . . . . One dollar per copy

No. 474. THE INFLUENCE OF DIFFERENT ACTIVITIES ON THE APPARENT LENGTH OF TIME

By JOHN C. LOEHLIN

No. 481. PHOTOSENSITIZATION IN RELATION TO MEAN AND STANDARD DEVIATION VALUES

By RICHARD T. ZEGERS

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Do you have problems?

Do you have problems in programming your experimental procedure? In presenting stimuli? In locating an appropriate reinforcer? In recording your data? The *Journal of the Experimental Analysis of Behavior* carries a regular series of instrumentation notes under the editorship of D. G. Anger. The titles of some recent notes in this series include:

> Devices for delaying the operation and release of relays—An instant-start circuit for fluorescent lamps—A transistorized circuit for recording contact responses—Graphic aids for the amateur draftsman—Mercury-wetted contact relays—Program tapes—A precision liquid feeding system controlled by licking behavior—A simple interval-programming circuit—Arc suppression in shock circuits—A continuous stencil duplicator—A speech rate meter for vocal behavior analysis—A multiple-unit pigeon apparatus—Vibration as a reinforcer for infant children—A semi-automatic teaching machine—A reliable, low-cost generator for audio-stimuli—Instantaneous lamp replacement—Some electronic control units for operant behavior studies: I. A response and reinforcement contingency translator; II. A random ratio generator; III. A timer and counter with print-out.

So that your students may also have access to these notes, make sure your library subscribes.

Journal of the Experimental Analysis of Behavior

Annual subscriptions, by the volume: Regular, $12.00; APA members, $9.50; Students, $6.50. All orders and business correspondence should be sent to Kay Dinsmoor, Journal of the Experimental Analysis of Behavior, Department of Psychology, Indiana University, Bloomington, Indiana.